MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo No. 613 June, 1981

A COMPUTATIONAL THEORY OF VISUAL SURFACE INTERPOLATION

W.E.L. Grimson

ABSTRACT: Computational theories of structure from motion [Ullman, 1979] and stereo vision [Marr and Poggio, 1979] only specify the computation of three-dimensional surface information at special points in the image. Yet, the visual perception is clearly of complete surfaces. In order to account for this, a computational theory of the interpolation of surfaces from visual information is presented.

The problem is constrained by the fact that the surface must agree with the information from stereo or motion correspondence, and **not** vary radically between these points. Using the image irradiance equation [Horn, 1977], an explicit form of this *surface consistency constraint* can be derived [Grimson, 1981c].

To determine which of two possible surfaces is more consistent with the surface consistency constraint, one must be able to compare the two surfaces. To do this, a functional from the space of functions to the real numbers is required. In this way, the surface most consistent with the visual information will be that which minimizes the functional. To ensure that the functional has a unique minimal surface, conditions on the form of the functional are derived. In particular, if the functional is a complete semi-norm which satisfies the parallelogram law, or the space of functions is a semi-Hilbert space and the functional is a semi-inner product, then there is a unique (to within an element of the null space of the functional) surface which is most consistent with the visual information.

It can be shown, based on the above conditions plus a condition of rotational symmetry, that there is a vector space of possible functionals which measure surface consistency, this vector space being spanned by the functional of quadratic variation and the functional of square Laplacian [Brady and Horn, 1981]. Arguments based on the null spaces of the respective functionals are used to justify the choice of the quadratic variation as the optimal functional.

Algorithms for computing the surface which minimizes quadratic variation in the case of exact surface interpolation and in the case of surface approximation are outlined and illustrated on a series of synthetic and actual surface interpolation examples.

This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's artificial intelligence research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-80-C-0505 and in part by National Science Foundation Grant MCS77-07569.

1. Introduction

Although our world has three spatial dimensions, the projection of light rays onto the retina presents our visual system with an image of the world that is inherently two-dimensional. We must use such images to physically interact with this three-dimensional world, even in situations new to us, or with objects unknown to us. That we do so easily implies that one of the functions of the human visual system is to reconstruct a three-dimensional representation of the world from its two-dimensional projection onto our eyes.

Methods that could be used to effect this three-dimensional reconstruction include stereo vision [Marr and Poggio, 1979; Grimson, 1980, 1981a; Mayhew and Frisby, 1981] and structure from motion [Ullman, 1979a]. Both of these methods may be considered as correspondence techniques, since they rely on establishing a correspondence between identical items in different images, and using the difference in projection of these items to determine surface shape. That is, correspondence methods compute surface information by:

(1) Identifying a location in the physical scene in one image;

(2) Identifying the corresponding location in a second image; either a second image taken from a different viewpoint in space (stereo) or a second image taken from a different viewpoint in time (structure from motion); and

(3) Computing a three-dimensional surface value, representing the distance of the point relative to some base point, based on the difference in the positions of the two corresponding points in the images.

Current computational theories of these processes [Marr and Poggio, 1979; Mayhew and Frisby, 1981; Ullman, 1979a] argue that the correspondence process cannot take place at all points in an image. Rather, the first stage of the correspondence process is to derive a symbolic description of points in the image at which the irradiance undergoes a significant change [Marr and Hildreth, 1980]. This symbolic representation (called the *primal sketch* [Marr, 1976; Marr and Hildreth, 1980]) forms the input to the second stage of the process in which the actual correspondence is computed. As a consequence of the form of the input, the correspondence process can compute explicit surface information only at scattered points in the image. Yet our perception is clearly of complete surfaces. (For example, in Figure 1, a sparse random dot stereogram yields the vivid perception of a square floating in space above a background plane, rather than a collection of dots suspended in space.) The problem

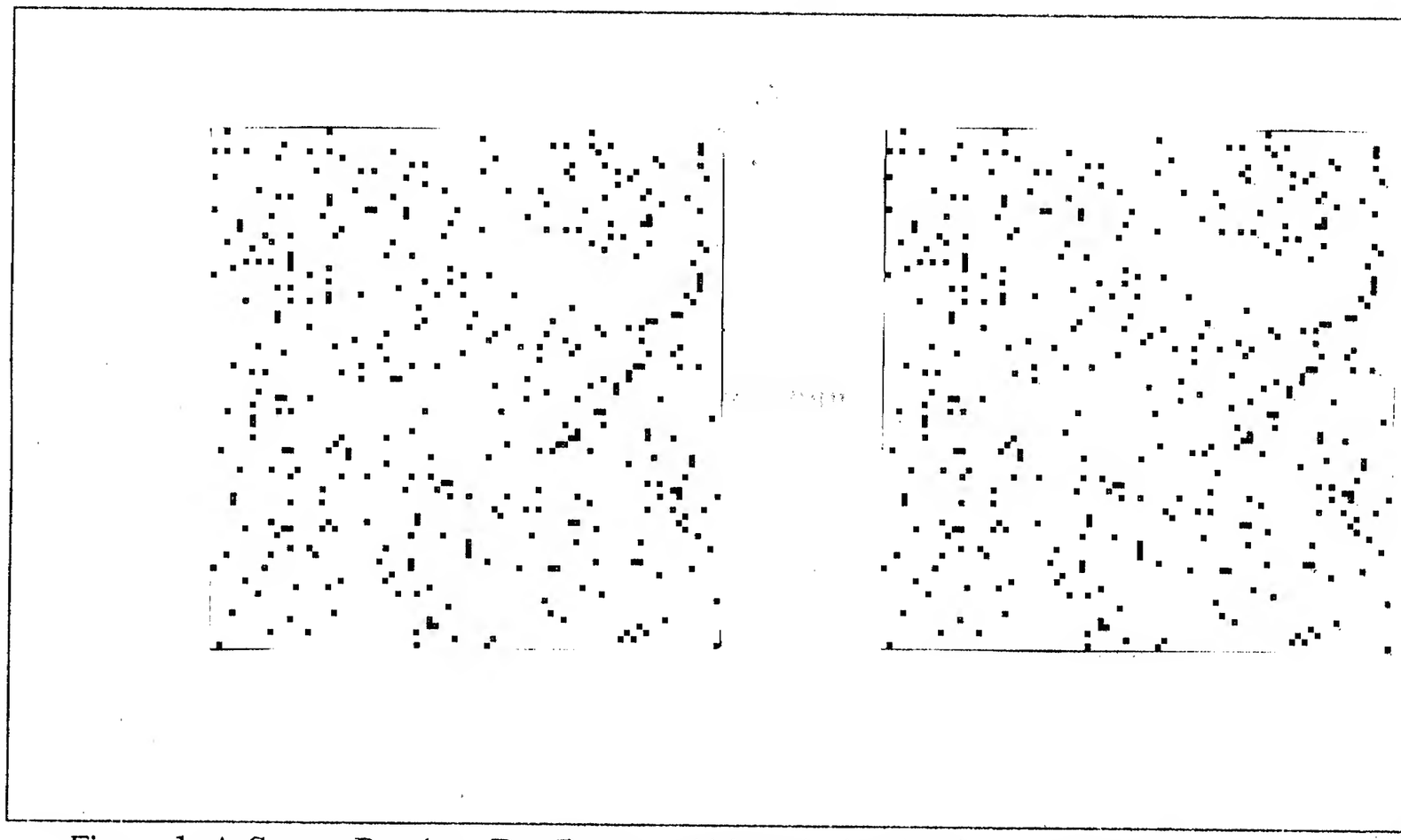

Figure 1. A Sparse Random Dot Pattern. Although the density of dots is very small, the perception obtained upon fusing this pattern is one of two disjoint planes, rather than dots isolated in depth.

to be addressed in this paper is that of computing complete surface representations, by interpolating an initial representation consisting of sparse surface values.

We will examine this surface interpolation problem at two levels. The first level is to consider the strictly mathematical question of surface reconstruction, independent of its relevance to the human visual system. Suppose one is given a visual process which determines surface information at points corresponding to relevant changes in the images. In general, there will be many possible surfaces consistent with these initial surface points. For example, consider the boundary conditions provided by a circular arc, along which the depth is constant. The possible surface consistent with these known points include a flat disc, a sphere and even the highly convoluted surface formed by a radial sine function (see Figure 3). How do we distinguish the correct one? Mathematically, we need to be able to compare two possible surfaces, in order to determine which is "better". This can be done by defining a functional Θ from the space of surfaces to the real numbers, so that comparing surfaces can be accomplished by comparing corresponding real numbers. Provided $\Theta(f) < \Theta(g)$ whenever surface f is "better" than surface g, the "best" surface to fit through the known points is that which minimizes Θ. There are two problems to solve here: (1) What does it mean for f to be "better" than g? and (2) Under what conditions does a unique "best" surface exist?

Once these questions have been answered and an appropriate functional has been derived, we can turn to the second level, which is to consider a specific algorithm for finding the surface that optimizes the functional. Because our intent is to consider models for the interpolation process as it occurs in the human visual system, we will restrict our attention to biologically feasible algorithms [Ullman, 1979b; Grimson, 1981b].

The motivation for considering the interpolation problem first mathematically, independent of the specifics of the human system, and then algorithmically, incorporating specific biological constraints, is based on the assumption that one can consider the visual system as a symbol manipulation process [Marr 1976, 1981; Marr and Poggio, 1977]. This implies that the meaning of the symbols being manipulated can be distinguished from the physical embodiment of those symbols. Hence, one can deal with the mathematical consideration of the information processing which is occurring, independent of the implementation of that processing (whether in transistors or neurons). The rationale for this view lies in the belief that any computational theory should address the fundamental questions of the information processing necessary to perform the task, and that such computational theories are independent, to a large extent, of the method used to compute them. The initial goal is thus to determine computational constraints on the interpolation problem, based on the input and output representations of the process, and based on the structure of the computation required to transform one representation into the other. Note that a computational theory of the information processing is applicable both to the human visual system, and to applications areas (such as high-altitude photomapping, hand-eye coordination systems, industrial robotics, and inspection of manufactured parts) where it is useful to create a complete specification of surface shape.

While we shall initially concentrate on the mathematical aspects of visual surface interpolation, the problem is not completely isolated from the human visual system. If we view the human early visual system as a symbolic manipulator, we can consider visual processing as a series of transformations from one representation to another [Marr, 1976, 1981]. In particular, three stages can be identified (see Figure 2). From the images, one transforms to a description, called the primal sketch, of those locations at which the image irradiances change. Next, primal sketch descriptions of several images are matched, either by the stereo or motion computation, to obtain a description of surface information at the zero-crossings. This representation is called the raw $2\frac{1}{2}$D sketch. Finally, the raw $2\frac{1}{2}$D sketch is interpolated to obtain complete surface descriptions, called the full $2\frac{1}{2}$D sketch [Marr,

1978; Marr and Nishihara, 1978]. The first two stages have been considered elsewhere [Marr, 1976; Marr and Hildreth, 1980; Hildreth, 1980; Marr and Poggio, 1979; Grimson, 1980, 1981a, 1981b; Ullman, 1979a]. It is the final stage — the problem of surface interpolation — that is considered here. to obtain complete surface descriptions, called the full $2\frac{1}{2}$D Sketch [Marr, 1978; Marr and Nishihara, 1978]. The first two stages have been considered elsewhere [Marr, 1976; Marr and Hildreth, 1980; Hildreth, 1980; Marr and Poggio, 1979; Grimson, 1980, 1981a, 1981b; Ullman, 1979a]. It is the final stage – the problem of surface interpolation – that is considered here.

The important point is that the form of the input and output representations can influence the design of the transformation. Here, we shall assume that the input representation consists of explicit surface information, such as distance or relative distance, along the zero-crossings of the convolved image (these terms will be given technical definitions in Section 2). The output representation will be a complete specification of surface information, where by complete, we mean that an explicit distance value should be computed at every point on some grid representation of the scene. Our main concern in this paper is with the computational constraints needed to transform the input representation into the output representation.

Although surface values at all points of the image are important, there is another aspect of surface information which should also be made explicit in the output representation. This is the set of discontinuities in surfaces; the occluding contours, both subjective and objective. Marr [1978] argues that the $2\frac{1}{2}$-D sketch should be a viewer-centered representation which includes both explicit surface information, such as depth and surface orientation, and explicit contours of surface discontinuities. In this paper, the concentration is on the problem of creating explicit surface information at all points of the surface. The question of surface discontinuities will be outlined, and possible algorithms suggested, but an implementation of this stage has not been completed.

2. Consequence of the Correspondence Problem

We indicated above that we would concentrate on correspondence methods which could effect the three-dimensional surface reconstruction; stereopsis [Marr, 1980; Marr and Poggio, 1979; Grimson, 1980, 1981a] and structure from motion [Ullman, 1979a]. The three main steps of the correspondence problem are: (1) identify a location in the physical scene in one image; (2) identify the corresponding location in a second image; either a second image taken from a different viewpoint

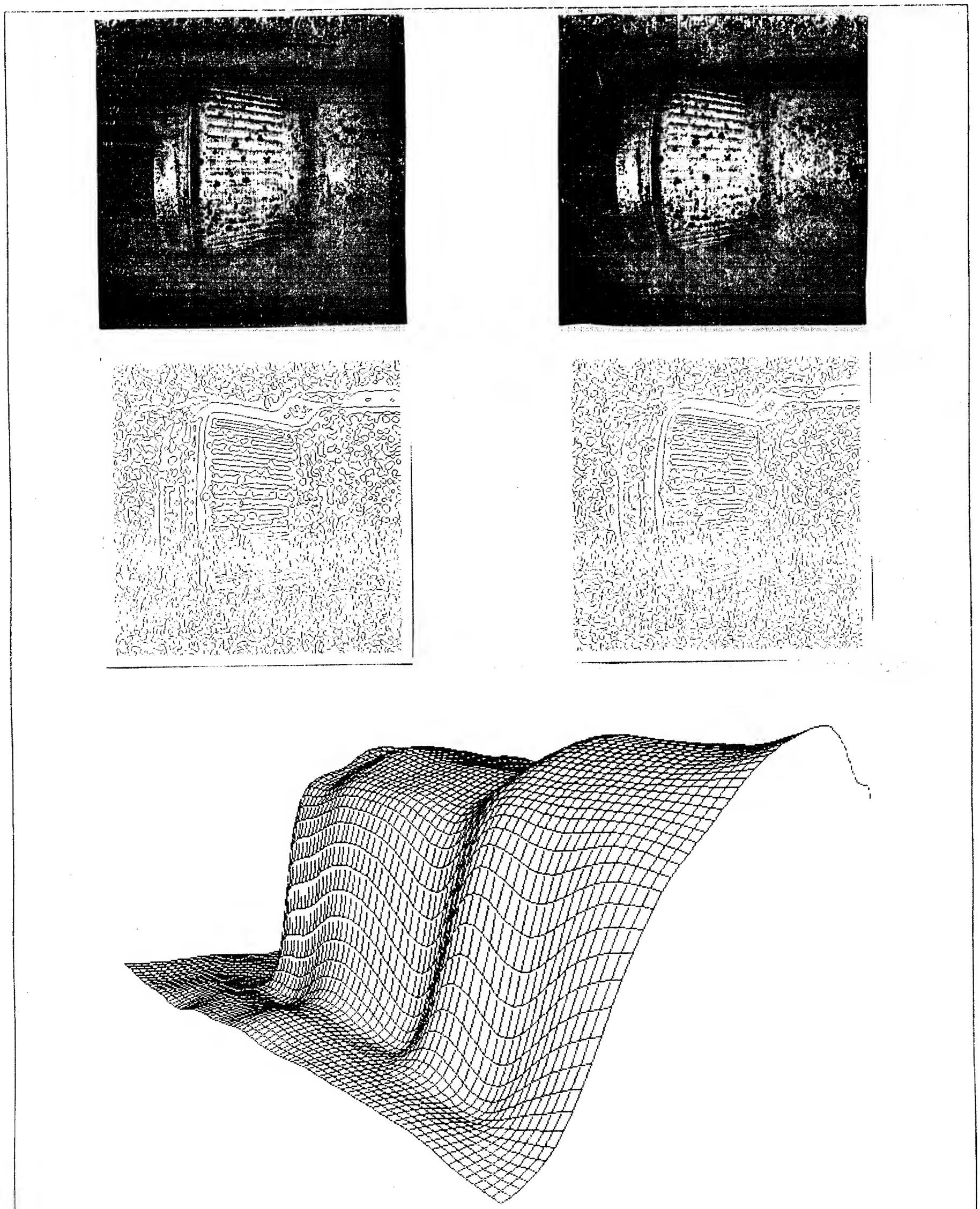

Figure 2 Example of Processing. The top pair of images is a stereo pair of a scene. The middle pair illustrates the zero-crossings obtained from the images for one size of $\nabla^2 G$. The final image illustrates an interpolated surface description, formed by matching the zero-crossing descriptions, computing the distance to those points based on the difference in projection, and then interpolating the result.

(stereo vision), or a second image taken at a later point in time (motion); and (3) compute a three-dimensional surface value, representing the distance of the point relative to some base point, based on the difference in the positions of the two corresponding points in the images.

If one can identify a location beyond doubt in the two images, then the correspondence problem is trivial. It can be demonstrated, however, that both the stereo computation and the motion computation can take place on very primitive descriptions of the images [Julesz, 1960; Ullman, 1979a]. As a consequence, the difficulty of the problem, for human vision, lies in the correspondence problem — which item in one image matches which item in the other. The reason for this is that for any primitive element from one description, there are liable to be many possible matching elements from the other description. This is especially true if image irradiance values are used as the basic descriptions. Consider an image of a matte-painted wall with uniform illumination. Given a small element of that wall from one image, it is virtually impossible to distinguish which small element from the other view matches it. On the other hand, if there is a scratch or texture marking on the wall, it is likely that such a location can be distinguished in the two views. This suggests that the representation upon which the correspondence operation takes place should reflect those positions in an image at which some physical property of the underlying surface is changing. This representation is called the primal sketch [Marr, 1976; Marr and Hildreth, 1980].

Marr and Hildreth [1980, also Hildreth, 1980] have refined the preceeding intuitive argument into more rigorous computational arguments, in conjunction with evidence from neurophysiology and psychophysics. They argue that the primal sketch representation is computed by determining those locations in an image at which the corresponding surface location undergoes a change in one of its physical properties, for example, reflectivity, surface orientation, texture or surface material. Such changes will correspond to a step change in image irradiance, at some scale. There are many ways of detecting the irradiance changes. Marr and Hildreth argue on various grounds for using the following scheme:

(1) Convolve the image with a set of filters given by the Laplacian applied to a Gaussian,

$$\nabla^2 G(r,\theta) = \left(\frac{r^2 - 2\sigma^2}{\sigma^4}\right) e^{\left(-\frac{r^2}{2\sigma^2}\right)}$$

where σ is a constant determined from psychophysical data.

(2) Locate all non-trivial *zero-crossings* in the convolved irradiances. A non-trivial zero-crossing is a point at which the convolved irradiance values change from positive to negative, or vice versa.

These zero-crossings form the basic representation upon which the later visual processing takes place.

Given this representation of the images, the correspondence problem can now be solved. Both Ullman's [1979a] theory of motion perception and Marr and Poggio's [1979] theory of stereo vision perform this computation on such primal sketch descriptions. As a consequence, explicit three-dimensional surface information (such as distance, or surface orientation) can be computed only at points corresponding to zero-crossings in the primal sketch. This would yield a sparse surface representation. Yet clearly, our perception is of complete surfaces (see for example Figure 1). In addition, a "nice" boundary is found for the central square. This implies that once the correspondence problem is solved, either by the stereo computation or by the motion computation, an interpolation must be performed between the surface values given at the zero-crossings, to obtain a complete surface description, and surface discontinuities should be explicitly demarked.

3. The Surface Consistency Constraint

We now turn to the heart of the matter, the computational constraints involved in the process of creating complete surface specifications, by interpolating between known points. We are given as basic input to the interpolation process, the zero-crossings of a convolved image, with depth information computed along these zero-crossing contours. Suppose one were to attempt to construct a complete surface description based only on the surface information known along the zero-crossings. An infinite number of surfaces would consistently fit the boundary conditions provided by these surface values. Yet there must be some way of deciding which surface, or at least which small family of surfaces, could give rise to the zero-crossing descriptions. This means that there must be some additional information available from the visual process which, when taken into account, will identify a class of nearly indistinguishable surfaces that represent the visible surfaces of the scene.

In order to determine what information is available from the visual process, one must first carefully consider the process by which the zero-crossing contours are generated. The Marr-Hildreth theory of edge detection [Marr and Hildreth, 1980; Hildreth, 1980] relies on the fact that sudden changes in the reflectance of a surface, for example, caused by surface scratches or texture markings, will give rise to zero-crossings in the convolved image. Sudden or sharp changes in orientation or

shape of the surface will under most circumstances also give rise to zero-crossings. This fact can be used to constrain the possible shapes of surfaces which could give rise to particular surface values along zero-crossing contours.

We illustrate the basic argument with an example. Suppose one is given a closed zero-crossing contour, within which there are no other zero-crossings. An example would be a circular contour, along which the depth is constant. There are many surfaces which could fit this set of boundary conditions (see Figure 3). One such surface is a flat disk. However, one could also fit other smooth surfaces to this set of boundary conditions. For example, the highly convoluted surface formed by $\sin\left(\sqrt{x^2+y^2}\right)$ would be consistent with the known disparity values. Yet in principle, such a rapidly varying surface should give rise to other zero-crossings. This follows from the observation that if the surface orientation undergoes a periodic variation, then it is likely that the irradiance values will also undergo such a variation. Since the only zero-crossings are at the borders of the object, this implies that the surface $\sin\left(\sqrt{x^2+y^2}\right)$ is not a valid representative surface for this set of boundary conditions.

Hence, the hypothesis is that the set of zero-crossing contours contains implicit information about the surface as well as explicit information. If one can determine a set of conditions on the surface shape that cause inflections in the irradiance values, then one may be able to determine a likely surface structure, given a set of boundary conditions along the zero-crossing contours.

3.1 No News is Good News

In general, any one of a multitude of widely varying surfaces could fit the boundary conditions imposed along the zero-crossings. To be completely consistent with the imaging process, however, such surfaces must meet both explicit conditions and implicit conditions. The explicit conditions are given by the depth or surface orientation values along the zero-crossing contours. The implicit conditions are that the surface must not impose any zero-crossing contours other than those which appear in the convolved image. This implicit condition leads to the *surface consistency constraint*, namely:

The absence of zero-crossings constrains the possible surface shapes.

Just as the presence of a zero-crossing tells us that some physical property is changing at a given location, the absence of a zero-crossing tells us the opposite, that no physical property is changing.

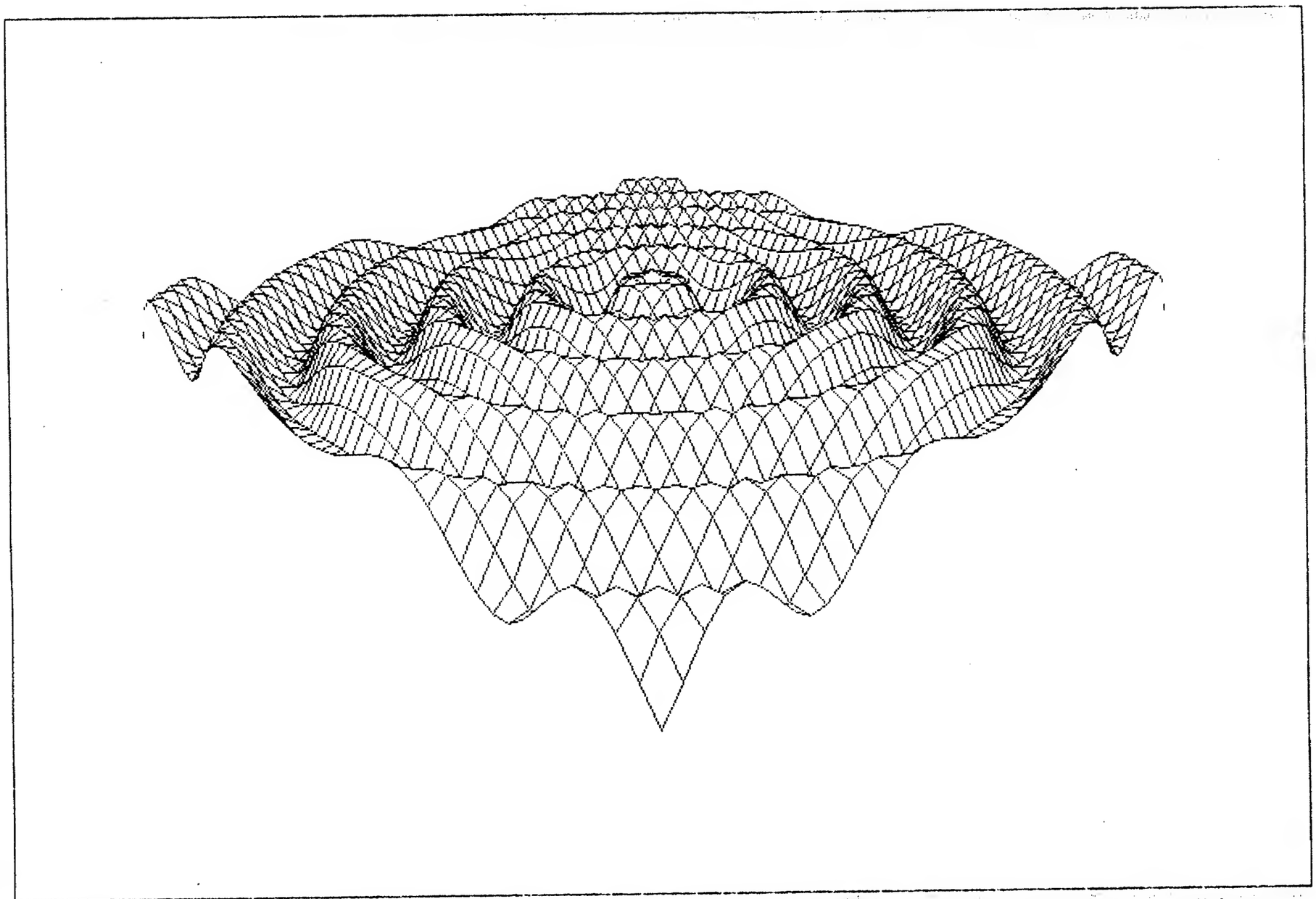

Figure 3. Possible Surfaces Fitting Depth Values at Zero-Crossings. Given boundary conditions of a circular zero-crossing contour, along which the depth is constant, there are many possible surfaces which could fit the known depth points. Two examples are a flat disk, and the highly convoluted surface formed by $\sin\left(\sqrt{x^2+y^2}\right)$, shown here.

and in particular that the surface topography is not changing in a radical manner. We informally refer to this constraint as *no news is good news*, since it says that the surface cannot change radically without informing us of this fact by means of zero-crossings.

In order to make explicit any constraints on the shape of the surface for locations in the image not associated with a zero-crossing, one must carefully examine the image formation process. Many factors are involved in the formation of image irradiances. As a consequence, changes in any of those factors can cause a change in the image irradiances, and hence a zero-crossing in the convolved image. For example, a change in surface material can correspond to a change in albedo, and hence to a zero-crossing in the convolved image; a discontinuity in depth can correspond to a change in the illumination striking the surface, and hence to a zero-crossing; and a discontinuity in surface orientation can correspond to a change in the amount of illumination reflected toward the viewer, and hence to a

zero-crossing. We are interested in showing that the inverse is also true — in particular, that in regions in which the illumination and albedo are roughly constant, the absence of a zero-crossing implies that the surface shape cannot be changing in a radical manner.

The basic result, corresponding to the intuitive argument of Figure 3, is that the probability of a zero-crossing occurring in regions where the illumination is roughly constant and the surface material does not change is a monotonic function of the variation in the orientation of the surface normal. (An analytic proof may be found in Grimson [1981c].) This provides a constraint on the possible surfaces that could be interpolated through a set of known points, and is referred to as the *surface consistency constraint.* It means that the probability of a zero-crossing increases as the variation in surface orientation increases. By inverting this argument, the best surface to fit the known data is that which minimizes the variation in surface orientation since this surface is most consistent with the zero-crossings in the convolved image.

4. The Computational Problem

We are now ready to consider the computational problem associated with the task of constructing complete surface specifications consistent with the information contained in the zero-crossings. The modules of early visual processing, such as stereo or structure-from-motion, provide explicit information about the shapes of the surfaces at specific locations in the images, corresponding to the zero-crossings of the convolved images. The surface consistency constraint indicates implicit information about the shapes of the surfaces between the zero-crossings, stating that between known depth values, the surface cannot change in a radical manner, since such changes would usually give rise to additional zero-crossings. These two factors will now be combined, to obtain a complete surface specification.

4.1 Using The Surface Consistency Constraint

Suppose we are given a set of known depth points. We want a method for finding a surface to fit through these points that is "most consistent" with the surface consistency constraint. We will find the most consistent surface in two ways. In the *surface interpolation* problem we construct a surface that exactly fits the set of known points. The problem can be relaxed somewhat into a *surface*

approximation problem, by only requiring that the surface approximately fit the known data and be smooth in some sense.

Given the initial boundary conditions of the known depth values along the zero-crossing contours, there is an infinite set of possible surfaces that fit through those points. We need to be able to compare members of this set of all possible surfaces fitting through those points, to determine which surface is more consistent. If we can do this, then the "most consistent" surface can be found by comparing all possible surfaces. A traditional method for comparing surfaces is to assign a real number to each surface. Then, in order to compare the surfaces, one need only compare the corresponding real numbers. The assignment of real numbers to possible surfaces is accomplished by defining a functional, mapping the space of possible surfaces into the real numbers, $\Theta: X \mapsto \Re$. This functional should be such that the more consistent the surface, the smaller the real number assigned to it. In order to satisfy the surface consistency constraint, the functional should measure variation in surface orientation. In this case, the most consistent surface will be the surface that is minimal under the functional. (For further details and background information about the use of functionals is, see, for example, Rudin [1973].)

The key mathematical difficulty is to guarantee the *existence* and *uniqueness* of a solution. In other words, we need to guarantee that there is at least one surface which minimizes the surface consistency constraint, and to guarantee that any other surface passing through the known points, for which the functional measure of surface consistency has the same value, is indistinguishable from the first surface. This issue is not just a mathematical nicety, however, but is essential to the solution of many computational problems. Suppose we devise an iterative algorithm to solve some problem. What happens if we cannot guarantee the existence of a solution? The iterative process could be set off to solve an equation and never converge to an answer — clearly an undesirable state. Further, suppose a solution exists but is not guaranteed to be unique. Then an iterative process might converge to one solution when started from one initial point, and converge to another solution when started from a different initial point. Although small variations in the different solutions might be acceptable, the solutions should not differ in a manner inconsistent with our intuition about the problem. Thus, in the case of visual surface interpolation, the real trick is to find a functional which accurately measures the variation in surface orientation, as well as guarantees the existence of a unique best surface (or a family of indistinguishable surfaces).

How can we guarantee the existence and uniqueness of a solution? In our particular case of surface interpolation, we will be using the calculus of variations to determine a system of equations which the most consistent surface must satisfy, by applying the calculus to the situation of fitting a thin plate through a set of known points. While this system of equations characterizes the minimal surface, it does not guarantee uniqueness. The form of the boundary conditions (the set of known points) will determine the size of the family of minimal surfaces. Unfortunately, determining the types of input for which a unique solution exists is generally very hard. Instead, we will exploit a general case of the mathematical existence of a solution with the weakest possible conditions on the functional. That is, we will determine a weak set of conditions on the functional that are needed to ensure that a unique most consistent surface, or at least a unique family of surfaces that are most consistent, will exist. We will show that if the functional is an inner-product on a Hilbert space of possible surfaces, then a unique most consistent surface will exist. (A Hilbert space is an extension of normal Euclidean space — basically a vector space in which a dot product operation exists, so that we can relate the vectors to the real line, and in which functions are usually used in place of the normal notion of vector.)

In general, it is extremely difficult to find a functional that measures surface consistency and satisfies the conditions of an inner-product. Hence, we will show that if the functional is a semi-inner product on a semi-Hilbert space of possible surfaces, then the most consistent surface is unique up to possibly an element of the null space of the functional. The null space is simply the set of surfaces that cannot be distinguished by the functional from the surface which is zero everywhere. In this way, the family of most consistent surfaces can be found. Based on the form of the null space, we can determine whether or not the differences in minimal surfaces are intuitively indistinguishable, and what conditions on the known boundary values will guarantee a unique minimal surface, from this family.

Having derived conditions on the functional, we need to show that there is such a functional. The surface consistency constraint implies that the functional should measure variation in surface orientation over an area of the surface. Although the condition of a semi-inner product is a mathematical requirement needed to guarantee a solution, it does not restrict in an unreasonable way the kinds of surfaces we consider, and gives rise to at least two very natural functionals, both of which can be derived from the calculus of variations: one measures the integral of the square Laplacian applied

to the surface and the other measures the quadratic variation of the local x and y components of the surface orientation. Besides the mathematical justifications, we will also show practical and intuitive reasons in support of such functionals.

Given that there are at least two possible functionals, are there others? We will show that if we require a functional that is:

1. a monotonic function of the variation in surface orientation,
2. a semi-inner product, and
3. rotationally symmetric,

then there is a vector space of possible functionals, spanned by the square Laplacian and the quadratic variation. In other words, there is a family of possible functionals, given by all linear combinations of these two basic functionals.

Given that there is more than one possible functional, how do they differ? Using the calculus of variations, and some results from mathematical physics, we will show that the surfaces which minimize these functionals will be roughly identical in the interior of a region and will differ only along the boundaries of a region. As well, the null spaces of the functionals will differ, implying different families of most consistent surfaces corresponding to each functional. We know that the minimal surface is unique up to possibly an element of the null space. Since we require that the solution surface either be unique, or a member of an indistinguishable family of solutions, the size of the null space is important in judging the value of a functional. Based on this, we will argue that the quadratic variation is to be preferred over the square Laplacian. If we require that the surface pass through the known points, we can show that the form of the stereo data will force a unique most consistent surface for the case of quadratic variation, while this is unlikely for functionals such as the square Laplacian.

Thus, we assert on mathematical grounds that the best functional is one which measures quadratic variation in surface orientation, and the most consistent surface to fit to the stereo data is that which passes through the known points and minimizes the quadratic variation. In a later section, we will see examples of the types of minimal surfaces obtained under quadratic variation and the square Laplacian. It will be seen that the mathematical distinction in size of null space has a practical consequence, as the types of surfaces which minimize the square Laplacian will be seen to be inconsistent with our intuitive notion of the best surface to fit to the known points, while the surface which

minimize the quadratic variation will be seen to be much more consistent with our intuitive notion of the best surface.

4.1.1 The Problem is Well-Defined

If surfaces are to be compared, by using a functional from the space of surfaces to the real numbers, with the purpose of finding the surface that best satisfies the surface consistency constraint, it is necessary to ensure that such a goal is attainable. What conditions on the form of the functional, or on the structure of the space of functions, are needed to guarantee the existence of such a "best" surface? One key constraint on the functional is given by the following theorem. The main point of the theorem is that in order to ensure that the problem is well-defined the functional should have the characteristics of a semi-norm.

Theorem 1: Suppose there exists a complete semi-norm Θ on a space of functions H, and that Θ satisfies the parallelogram law (for definition, see proof of theorem). Then, every nonempty closed convex set $E \subset H$ contains a unique element v of minimal norm, up to an element of the null space. Thus, the family of minimal functions is

$$\{v + s \mid \ s \in S\}$$

where

$$S = \{v - w \mid \ w \in E\} \cap \mathcal{N}$$

and $\mathcal{N}$ is the null space of the functional

$$\mathcal{N} = \{u \mid \ \Theta(u) = 0\}.$$

Proof: [See for example Rudin, 1973]. Any space with a semi-norm defined on it can be associated with an equivalent normed space. Let W be a subspace of a vector space H. For every $v \in H$, let $\pi(v)$ be the coset of W that contains v,

$$\pi(v) = \{v + u: u \in W\}.$$

These cosets are elements of a vector space H/W called the quotient space of H modulo W. In this space, addition is defined by

$$\pi(v) + \pi(w) = \pi(v + w)$$

and scalar multiplication is defined by

$$\alpha\pi(v) = \pi(\alpha v).$$

The origin of the space H/W is $\pi(0) = W$. Thus, π is a linear map of H onto H/W with W as its null space.

Now consider the semi-norm Θ on the vector space H. Let

$$\mathcal{N} = \{v: \Theta(v) = 0\}.$$

This can easily be shown to be a subspace of H. Let π be the quotient map from H onto $H/\mathcal{N}$, and define a mapping Θ':$H/\mathcal{N} \mapsto \Re$,

$$\Theta'(\pi(v)) = \Theta(v).$$

If $\pi(v) = \pi(w)$ then $\Theta(v-w) = 0$. Since $|\Theta(v) - \Theta(w)| \leq \Theta(v-w)$, then $\Theta'(\pi(v)) = \Theta'(\pi(w))$ and Θ' is well defined on $H/\mathcal{N}$. It is straightforward to show that Θ' is a norm on $H/\mathcal{N}$.

Now we can prove the statement of the theorem. The set E, a subset of H, can be transformed into a set E' in the quotient space $H/\mathcal{N}$ while preserving the convexity and closure properties.

The parallelogram law states

$$[\Theta'(v+w)]^2 + [\Theta'(v-w)]^2 = 2[\Theta'(v)]^2 + 2[\Theta'(w)]^2.$$

Let

$$d = \inf\{\Theta'(v): v \in E'\}.$$

Choose a sequence $v_n \in E'$ such that $\Theta'(v_n) \mapsto d$. By the convexity of E', we know that $\frac{1}{2}(v_n + v_m) \in E'$ and so $[\Theta'(v_n + v_m)]^2 \geq 4d^2$. If v and w are replaced in the definition of the parallelogram law by v_n and v_m, then the right hand side tends to $4d^2$. But $[\Theta'(v_n + v_m)]^2 \geq 4d^2$, so one must have $[\Theta'(v_n - v_m)]^2 \mapsto 0$ to preserve the equality. Thus, $\{v_n\}$ is a Cauchy sequence in $H/\mathcal{N}$. Since the norm is complete, the sequence must converge to some $v \in E'$, with $\Theta'(v) = d$.

To prove the uniqueness, if $v, w \in E'$ and $\Theta'(v) = d, \Theta'(w) = d$ then the sequence $\{v, w, v, w, \ldots\}$ must converge, as we just saw. Thus $v = w$ and the element is unique.

We have proven that under the norm Θ' on the quotient space $H/\mathcal{N}$, the set E' has a unique minimal element. Hence, the structure of the quotient space implies that under the semi-norm Θ on

the space H, the set E has a unique minimal element v, up to possibly an element of the null space $\mathcal{N}$. In other words, the family of minimal elements is

$$\{v + s \mid \ s \in S\}$$

where

$$S = \{v - w \mid \ w \in E\} \cap \mathcal{N}.$$

∎

This theorem specifies one set of mathematical criteria needed to ensure that there exists a unique minimal element. Thus, if the surface consistency constraint could be specified by a functional that satisfied the conditions of a complete semi-norm, obeying the parallelogram law, it might be possible to show that there is a unique coset of "most consistent" surfaces. We would really prefer to be guaranteed a unique surface, rather than some set of surfaces. One way to tighten the result of the theorem is to require that the functional is a norm.

Corollary 1.1: If Θ is a complete norm on a space of functions H, which satisfies the parallelogram law, then every nonempty closed convex set $E \subset H$ contains a unique element v of minimal norm.

Proof: If the functional is a norm, the null space is the trivial null space, and the result holds uniquely. ∎

The theorem can be rephrased in terms of the surface interpolation problem as follows.

Corollary 1.2: Let the set of known points be given by

$$\{(x_i, y_i) \mid i = 1, \ldots, N\}$$

where the associated depth value is F_i. Let $\mathcal{F}$ be a vector space of "possible" functions on $\Re^2$ and let

$$U = \{f \in \mathcal{F} \mid f(x_i, y_i) = F_i \quad i = 1, \ldots, N\}$$

so that U is the set of functions that interpolate the known data $\{F_i\}$. Let Θ be a semi-norm, which measures the "consistency" of a function $f \in X$, that is, we shall say that f is "better" than g if $\Theta(f) < \Theta(g)$. If Θ is a complete semi-norm and satisfies the parallelogram law, then there exists a unique (to within a function of the null space of Θ) function $s \in U$ that is least inconsistent and interpolates the data. Hence the interpolation problem is well-defined.

Proof: Clearly U is a convex set since for any $f, g \in U$,

$$[\lambda f + (1 - \lambda)g](x_i, y_i) = (\lambda + 1 - \lambda)F_i = F_i$$

for any data point (x_i, y_i). Furthermore, U is closed, since if $f_n \in U$ and $f_n \mapsto f$, then $f(x_i, y_i) = F_i$ and $f \in U$. Then the previous corollary states that U has a unique (to within an element of the null space) element of minimal norm, which is exactly the desired "most consistent" surface. ∎

This corollary is a translation of Theorem 1 into the problem of interest to us, finding the surface most consistent with the known data from the stereo algorithm. It specifies a set of conditions under which the interpolation problem is well-defined. Here, the notion of well-defined refers to finding a solution to the interpolation problem that is unique, and by unique we mean up to possibly an element of the null space of the semi-norm. As a consequence, the extent and structure of the null space of any semi-norm chosen to incorporate the surface consistency constraint will be important in determining the utility of that semi-norm. In general, the smaller the null space, the tighter the constraint on the family of possible surfaces which can interpolate the known data. For example, if the possible variations in the least inconsistent surface were very large (due to the semi-norm having a large null space), then the utility of this semi-norm would have to be questioned. On the other hand, if the null space is small, and the resulting variations in possible least inconsistent surfaces were small, the semi-norm would have to be given serious consideration. We will see examples of possible functionals and their null spaces in Section 4.1.4.

Thus, Theorem 1 and Corollary 1.1 specify two different sets of sufficient, but not necessary, criteria for ensuring differing types of uniqueness. In both cases, the criteria applied directly to the structure of the functional. Of course, the real trick is to find a functional Θ which captures our intuition of variation in surface orientation and meets the requirements needed to guarantee a unique solution.

4.1.2 The Space of Functions

Theorem 1 describes a set of sufficient conditions for obtaining a unique family of minimal surfaces. The fundamental point is that we require a complete parallelogram semi-norm to ensure a unique solution. These conditions precisely define a semi-inner product, and hence the space of functions over which we seek a minimum must be a semi-Hilbert space.

Corollary 1.3: If $\mathcal{F}$ is a semi-Hilbert space of possible surfaces, and $\Theta(v) = \mu(v, v)^{\frac{1}{2}}$ is an inner product semi-norm, where $\mu(v, v)^{\frac{1}{2}}$ is the semi-inner product on the space $\mathcal{F}$, then there exists a unique surface in $\mathcal{F}$ (possibly to within an element of the null space of the semi-norm) that minimizes the semi-norm Θ over all surfaces.

Proof: By the definition of Hilbert space, the semi-norm is complete. It is easy to show that it satisfies the parallelogram law from the definition of $\Theta(v) = \mu(v, v)^{\frac{1}{2}}$. Thus, if the space of functions is a semi-Hilbert space, then, by Theorem 1, the interpolation problem is guaranteed to have a unique minimal solution, possibly to within an element of the null space. ∎

Corollary 1.4: If $\mathcal{F}$ is a Hilbert space of possible surfaces, and $\Theta(v) = \mu(v, v)^{\frac{1}{2}}$ is an inner product norm, where $\mu(v, v)^{\frac{1}{2}}$ is the inner product on the space $\mathcal{F}$, then there exists a unique surface in $\mathcal{F}$ that minimizes the norm Θ over all surfaces. ∎

4.1.3 The Form of the Functional

The major problem is to determine the functional Θ. The surface consistency constraint related the consistency of a surface to the variation in surface orientation. This forms the first constraint on the functional. Theorem 1 states that if the functional is a complete, parallelogram, semi-norm, then the problem has a well-defined solution. This forms the second constraint on the functional. Thus, if a functional can be found that is a complete, parallelogram semi-norm, and that is a monotonic function of the variation in surface orientation, then this functional will constitute an acceptable measure of surface inconsistency. Any surface that is minimal under such a functional would then be an acceptable reconstruction of the original surface in space.

The problem may be considered in the following manner. With every point on the surface f, associate a pair of partial derivatives, $f_x = p, f_y = q$, and hence, an orientation. Each point on the surface may be mapped to a point in a space spanned by p and q axes, the gradient space [Huffman, 1971; Mackworth, 1973; Horn, 1977]. With each surface patch, one may then associate a neighborhood of p-q space by mapping the p and q values associated with each point on the surface into gradient space. This neighborhood will be referred to as the p-q neighborhood spanned by the surface patch.

The surface consistency constraint implies that the probability of a zero-crossing occuring within a patch of the surface is monotonically related to the size of the p-q neighborhood spanned by the

surface patch. The surface consistency theorem [Grimson, 1981c] embodies a specific method for measuring this probability. Any functional that is monotonically related to the size of the p-q neighborhood will suffice, however. This is important, since it is also necessary that the functional be a complete, parallelogram semi-norm. Thus, the problem reduces to specifiying such a complete, parallelogram semi-norm, which monotonically measures the surface consistency constraint by measuring a monotonic function of the size of the p-q neighborhood spanned by each surface patch. To find the most consistent surface, one need only find the surface that minimizes this functional over each patch. Note that the surface will be "most consistent" only relative to the functional. There may be several functionals that are complete, parallelogram, semi-norms and that are monotonic functions of the variation in surface orientation. Each will give rise to slightly different minimal surfaces.

Of course, there are some constraints on the minimization of the measure of the p-q neighborhood. For example, each surface patch cannot be minimized in isolation. To see this, consider a cylindrical (or one-dimensional) surface. Between any two adjacent zero-crossing points, the minimization of the variation in surface orientation (related to the size of the neighborhood in p-q space) would result in a single point in gradient space, corresponding to a planar surface between the known depth values. The problem with this simple method of reducing surface inconsistency is that it does not account for the interaction of surface patches. In particular, such a method would result in a piecewise planar surface approximation. For any three consecutive zero-crossing points, such a method would introduce a discontinuity in surface orientation at the middle zero-crossing. This is unacceptable since the surface is required to be twice differentiable. Thus, there are some constraints on the manner in which the neighborhoods in p-q space are minimized.

We are thus faced with the following problem. We know from the boundary conditions provided by the stereo algorithm that the surface must pass through a set of known depth points located at the zero-crossings of the convolved images. We further know that at all other points in the image, the surface cannot change in such a manner as to cause an additional zero-crossing. With each surface portion, we can associate a measure of the probability of that surface implying a zero-crossing in the convolved image intensities. Since between zero-crossings, there are no other zero-crossings, this gives a measure of the inconsistency of this particular surface portion. To choose the least inconsistent surface, we must reduce this probability, as measured over all portions of the surface.

4.1.4 Possible Functionals

In this section, possible functionals Θ are considered, seeking complete, parallelogram semi-inner products where possible, since this will guarantee that the solution is unique to within the null space. However, it is important to stress that there may be several viable alternatives. The computational theory argued that the functional should measure the "consistency" of the surface. The attempt here is to define a measure based on this. The measure should be in a form that allows the constraints on the problem to be easily expressed. Also, the measure should be a semi-inner product on a semi-Hilbert space, in order to ensure a unique solution. We begin by considering several candidates in light of these requirements.

Case 1: One Dimension

For ease of discussion, the case of a cylindrical or one-dimensional surface will be considered first. By a cylindrical (or developable) surface we mean a second differentiable surface oriented along the y axis such that $\partial f/\partial y = q = c$, for some constant c.

Example 1.1: The variation in the normal to the curve is clearly related to its curvature. One could thus consider using a functional that directly measures curvature and integrates this measure over an area of the surface. To ensure that the functional is positive, this suggests using a functional of the form:

$$\Theta_1(f) = \left\{ \int \kappa^2 \, ds \right\}^{\frac{1}{2}} = \left\{ \int \frac{f_{xx}^2}{(1+f_x^2)^{\frac{5}{2}}} \, dx \right\}^{\frac{1}{2}},$$

where κ is the curvature of the curve at a point. (Recall that the subscript here implies a partial derivative, so that $f_{xx}^2 = (\partial^2 f/\partial x^2)^2$.) Although this is perhaps the most "natural" definition of a functional, it is not a semi-norm, and hence it is considered to be unacceptable. To see why it is not a semi-norm, consider the following. If f is in the space of surfaces, then

$$\begin{aligned}
\Theta_1(\alpha f) &= \left\{ \int \frac{\alpha^2 f_{xx}^2}{(1+\alpha^2 f_x^2)^3} \, dx \right\}^{\frac{1}{2}} \\
&= |\alpha| \left\{ \int \frac{f_{xx}^2}{(1+\alpha^2 f_x^2)^3} \, dx \right\}^{\frac{1}{2}} \\
&\neq \alpha \Theta_1(f).
\end{aligned}$$

This condition will be true only if $f_x \equiv 0$. While this is certainly far too restrictive a condition to place on the possible surfaces, it does suggest a possible alternative.

Example 1.2: A second choice is the quadratic variation of the gradient, which may be measured by:

$$\Theta_2(f) = \left\{ \int f_{xx}^2 \, dx \right\}^{\frac{1}{2}}$$

Note that it is a close approximation to the curvature of the curve, provided that the gradient f_x is small. Θ_2 is a semi-norm, so the surface that minimizes this norm will be unique to within an element of the null space of the semi-norm. The null space of Θ_2 is the set of all linear functions:

$$\mathcal{N} = \text{span}\{1, x\}$$

where

$$\text{span}\{v_1, \ldots, v_m\} = \{a_1 v_1 + \ldots + a_m v_m \mid a_1, \ldots, a_m \in \Re\}.$$

Not only does this form of the functional satisfy the mathematical criteria of a complete, parallelogram semi-norm, it has a strong relationship to the "natural" form $\Theta_1(f)$, since the restriction of small f_x is acceptable. Those cases in which f_x is not negligible correspond to situations in which the surface is rapidly curving away from the viewer. These situations are such that the curvature of the surface will cause it to be invisible in one eye — giving rise to occluding boundaries. In general, we can assume that the image does not consist solely of occluding boundaries, so that such occurrences will be rare in an image. Moreover, between such points, the surface will satisfy the restriction and the above semi-norm is well-suited to the interpolation problem.

Case 2: Two Dimensions

To each of the examples of the one-dimensional case, there is an analogous two-dimensional case.

Example 2.1: As in the one-dimensional case, one possibility is to measure the curvature of the surface. The curvature of a surface is usually measured in one of two ways.

For any point on the surface, consider the intersection of the surface with a plane containing the normal to the surface at that point. This intersection defines a curve, and the curvature of that curve can be measured as the arc-rate of rotation of its tangent. For any point, there are infinitely many normal sections, each defining a curve. As the normal section is rotated through 2π radians,

all possible normal sections will be observed. There are two sections of particular interest, that which has the maximum curvature and that which has the minimum. It can be shown that the directions of the normal sections corresponding to these sections are orthogonal. These directions are the *principal directions* and the curvatures of the normal sections in these directions are the *principal curvatures*, denoted κ_a and κ_b. It can be shown that the curvature of any other normal section is defined by the principal curvatures.

There are two standard methods for describing the curvature of the surface, in terms of the principal curvatures. One is the first (or mean) curvature of the surface

$$J = \kappa_a + \kappa_b.$$

The other is the second or Gaussian curvature of the surface

$$K = \kappa_a \cdot \kappa_b.$$

For a surface defined by the vector $\{x, y, f(x, y)\}$, these curvatures are given by

$$J = \frac{\partial}{\partial x}\left(\frac{f_x}{\sqrt{1+f_x^2+f_y^2}}\right) + \frac{\partial}{\partial y}\left(\frac{f_y}{\sqrt{1+f_x^2+f_y^2}}\right)$$

and

$$K = \frac{f_{xx}f_{yy} - f_{xy}^2}{\left(1+f_x^2+f_y^2\right)^2}.$$

Thus, there are two possibilities for the functional. One is to measure the first (or mean) curvature of the surface,

$$\Theta_3(f) = \left\{\int\int J^2\,dxdy\right\}^{\frac{1}{2}}$$

$$= \left\{\int\int \frac{\left(f_{xx}(1+f_y^2) + f_{yy}(1+f_x^2) - 2f_xf_yf_{xy}\right)^2}{\left(1+f_x^2+f_y^2\right)^3}\,dxdy\right\}^{\frac{1}{2}}.$$

As in the one-dimensional case, this is not a semi-norm, since

$$\Theta_3(af) = |a|\left\{\int\int \frac{\left(f_{xx}(1+a^2f_y^2) + f_{yy}(1+a^2f_x^2) - 2af_xaf_yf_{xy}\right)^2}{\left(1+a^2f_x^2+a^2f_y^2\right)^3}\,dxdy\right\}^{\frac{1}{2}}$$

$$\neq |a|\Theta_3(f).$$

However, if f_x and f_y are assumed to be small, then it is closely approximated by a semi-norm. In this case, consider

$$\Theta_4(f) = \left\{ \int \int \left(\nabla^2 f \right)^2 dxdy \right\}^{\frac{1}{2}}$$

This is a semi-norm, with null space consisting of all harmonic functions.

A second possibility for reducing curvature is to reduce the second or Gaussian curvature,

$$\Theta_5(f) = \left\{ \int \int K^2 \, dxdy \right\}^{\frac{1}{2}}.$$

By an argument similar to the above, it can be shown that this is not a semi-norm. Note that by using the above approximation of small f_x and f_y, we get the functional

$$\Theta_6(f) = \left\{ \int \int f_{xx} f_{yy} - f_{xy}^2 \, dxdy \right\}^{\frac{1}{2}}.$$

We will return to this form later.

Example 2.2: As in the one-dimensional case, one can also consider the quadratic variation. The quadratic variation in $p = f_x$ is given by

$$\int \int \left(p_x^2 + p_y^2 \right) dxdy$$

and the quadratic variation in $q = f_y$ is given by

$$\int \int \left(q_x^2 + q_y^2 \right) dxdy .$$

If the surface is twice continuously differentiable, then $p_y = q_x$, and by combining these two variations, one obtains the quadratic variation:

$$\Theta_7(f) = \left\{ \int \int \left(f_{xx}^2 + 2 f_{xy}^2 + f_{yy}^2 \right) dxdy \right\}^{\frac{1}{2}}.$$

Again, as in the one-dimensional case, this is a complete semi-norm that satisfies the parallelogram law. Hence, the space of interpolation functions has an element of minimal norm, which is unique up to an element of the null space, where the null space is the set of all linear functions:

$$\mathcal{N} = \text{span}\{1, x, y\}.$$

Duchon (1975, 1976) refers to the surfaces that minimize this expression as *thin plate splines* since the expression Θ_7 relates to the energy in a thin plate forced to interpolate the data.

4.2 Where Do We Stand?

We have seen that for the general surface interpolation problem, there are two constraints on the possible functionals. One is that the functional must measure a monotonic function of the variation in surface orientation. The other is that the functional should satisfy the conditions of a complete parallelogram semi-norm, or equivalently, a semi-inner product. If the functional satisfies these conditions, then we know that there will be a unique family of surfaces that minimize this functional and hence form a family of best possible surfaces to fit through the known information. In the examples sketched above we saw that there are at least two possible candidates for this functional, namely the square Laplacian,

$$\Theta_4(f) = \left\{ \int \int \left(\nabla^2 f \right)^2 dx dy \right\}^{\frac{1}{2}}$$

and the quadratic variation,

$$\Theta_7(f) = \left\{ \int \int \left(f_{xx}^2 + 2f_{xy}^2 + f_{yy}^2 \right) dx dy \right\}^{\frac{1}{2}}.$$

There are several points still to consider. Are there other possible functionals? How do the minimal solutions to these functionals differ? What criteria can be applied to determine which functional is best suited to our surface interpolation problem? What is the best functional under those criteria? In the remainder of this chapter, we shall consider these questions in detail. The point we shall develop is that the appropriate functional to apply is the quadratic variation, and thus the surface that minimizes this functional is most consistent with the imaging information.

4.3 Are There Other Functionals?

We have determined at least two functionals that meet our conditions. Are there other possible functionals, and if so, how do their minimal solutions differ from those of the square Laplacian and the quadratic variation?

To answer this question, we rely on a result of Brady and Horn [1981], which we sketch below. Recall that the basic conditions on the functional were that it measure a monotonic function of the variation in surface orientation, and that it be a semi-inner product. The first requirement suggests that the functional must involve terms that are functions of the second order partial derivatives of the

surface, since such terms will be related to the variation in surface orientation. The second requirement is needed to ensure the uniqueness of the solution. Recall that $\mu(f, g)$ is a semi-inner product if

1. $\mu(f, g) = \mu(g, f)$,
2. $\mu(f + g, h) = \mu(f, h) + \mu(g, h)$,
3. $\mu(\alpha f, g) = \alpha\mu(f, g)$,
4. $\mu(f, f) \geq 0$,

and that given a semi-inner product $\mu(f, g)$, one can define the desired functional by $\Theta(f) = \mu(f, f)^{\frac{1}{2}}$.

The difficult condition to satisfy is (3), which implies that the semi-inner product should not contain any constant terms. The conditions taken together imply that we should consider any quadratic form as a possible semi-inner product:

$$\mu(f, g) = \int\int \alpha f_{xx}g_{xx} + \beta f_{xy}g_{xy} + \gamma f_{yy}g_{yy} + \\ + \delta(f_{xx}g_{xy} + f_{xy}g_{xx}) + \epsilon(f_{xx}g_{yy} + f_{yy}g_{xx}) + \varsigma(f_{xy}g_{yy} + f_{yy}g_{xy}).$$

Thus, the corresponding functional will have the quadratic form:

$$\Theta(f) = \int\int \alpha f_{xx}^2 + \beta f_{xy}^2 + \gamma f_{yy}^2 + 2\delta f_{xx}f_{xy} + 2\epsilon f_{xx}f_{yy} + 2\varsigma f_{xy}f_{yy}.$$

The final condition we apply to the functional is that it be rotationally symmetric. This follows from the observation that if the input is rotated, the surface that fits the known data should not change in form, other than also being rotated.

Minimizing the quadratic form of the functional $\Theta(f)$ can be considered as finding the minimum over the integral of the function $(\Delta f)^T M \Delta f$ where Δf is the vector:

$$\Delta f = \begin{pmatrix} f_{xx} \\ f_{xy} \\ f_{yy} \end{pmatrix}$$

and M is the symmetric matrix

$$M = \begin{bmatrix} \alpha & \delta & \epsilon \\ \delta & \beta & \varsigma \\ \epsilon & \varsigma & \gamma \end{bmatrix}.$$

If R is a rotation matrix, then the condition of rotational symmetry is given by

$$(R\Delta f)^T M(R\Delta f) = \Delta f^T M \Delta f.$$

Vector algebra implies that we must have

$$R^T M R = M \qquad \text{or} \qquad R^T M = M R^{-1}.$$

Equating elements shows that the matrix M must have the form

$$M = \begin{bmatrix} \frac{\beta}{2} + \epsilon & 0 & \epsilon \\ 0 & \beta & 0 \\ \epsilon & 0 & \frac{\beta}{2} + \epsilon \end{bmatrix}.$$

There are two important consequences of this fact. The first is that the set of all possible functionals forms a vector space, since if M_1 and M_2 satisify the conditions, then so does $\sigma M_1 + \nu M_2$. The second is that this vector space of operators is spanned by the square Laplacian and the quadratic variation since:

$$\begin{bmatrix} \frac{\beta}{2} + \epsilon & 0 & \epsilon \\ 0 & \beta & 0 \\ \epsilon & 0 & \frac{\beta}{2} + \epsilon \end{bmatrix} = \epsilon \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix} + \frac{\beta}{2} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The first term of the sum corresponds to the square Laplacian while the second corresponds to the quadratic variation. Thus, for $\epsilon = 1$ and $\beta = 0$, the functional reduces to square Laplacian. For $\epsilon = 0$ and $\beta = 2$, the functional reduces to quadratic variation. Finally, if $\epsilon = \frac{1}{2}$ and $\beta = -1$ we obtain a functional which corresponds to the approximation to the integral of square Gaussian curvature derived in Section 4.1.4.

Thus, we have answered our second question. There are other possible functionals, but they are all linear combinations of the two basic functionals, the square Laplacian and the quadratic variation.

4.4 How Do the Functionals Differ?

Given that there are many possible functionals, all linear combinations of the square Laplacian, Θ_4, and the quadratic variation, Θ_7, we must consider how the solutions to the square Laplacian and the quadratic variation differ. In other words, is there any noticeable difference in the surfaces that minimizes these two functionals, subject to fitting through the stereo data? To answer this question, we shall rely on the Calculus of Variations, (see, for example, Courant and Hilbert [1953] and Forsyth [1960]). The salient points are outlined below.

4.4.1 Calculus of Variations

The calculus of variations is frequently used to solve problems of mathematical physics, and is applicable to our surface interpolation problem. In particular, we can use the calculus of variations to formulate differential equations associated with problems of minimum energy. Suppose we are given a thin elastic plate, whose equilibrium position is a plane, and whose potential energy under deformation is given by an integral of the quadratic form in the principal curvatures of the plate. We can consider the interpolation problem as one of determining the surface formed by fitting a thin elastic plate over a region $\mathcal{R}$ (with boundary Γ) and through the known points. Using a small deflection approximation, the potential energy is given by

$$\int\int_{\mathcal{R}} \left[(\nabla^2 f)^2 - 2(1-\mu)(f_{xx}f_{yy} - f_{xy}^2) \right] dxdy.$$

The solution to the interpolation problem is then the surface which has the minimum potential energy.

The calculus of variations can be used to characterize this problem by providing a set of differential equations (called the Euler equations) which the solution surface must satisfy. It can be shown (see Courant and Hilbert, [1953, p.251]) that the Euler equations for the interior of any region $\mathcal{R}$ are given by

$$\nabla^4 f = f_{xxxx} + 2f_{xxyy} + f_{yyyy} = 0,$$

except at the known points. Along the boundary contour Γ of the region, the solution surface must satisfy the equations (called the *natural boundary conditions*):

$$M(f) = -\nabla^2 f + (1-\mu)(f_{xx}x_s^2 + 2f_{xy}x_s y_s + f_{yy}y_s^2) = 0$$

$$P(f) = \frac{\partial}{\partial n}\nabla^2 f + (1-\mu)\frac{\partial}{\partial s}\left(f_{xx}x_n x_s + f_{xy}(x_n y_s + x_s y_n) + f_{yy}y_n y_s \right) = 0,$$

where $\partial/\partial n$ is a derivative normal to the boundary contour, $\partial/\partial s$ is a derivative with respect to arclength along the boundary contour and x_s, y_s and x_n, y_n are the direction cosines of the tangent vector and the outward normal respectively. The constant μ denotes the tension factor associated with the elastic material of the plate.

There are two subcases of particular interest. In the first case, suppose that the tension factor is given by $\mu = 1$. The energy equation reduces to

$$\int\int_{\mathcal{R}} (\nabla^2 f)^2 \, dxdy$$

which is simply the square Laplacian condition derived previously. The Euler equation is the biharmonic equation $\nabla^4 f = 0$ while the natural boundary conditions are

$$\nabla^2 f = 0$$
$$\frac{\partial}{\partial n}\nabla^2 f = 0,$$

along the boundary contour Γ. In the second case, suppose that the tension factor is given by $\mu = 0$. The energy equation reduces to

$$\int\int_{\mathcal{R}} (f_{xx}^2 + 2f_{xy}^2 + f_{yy}^2)\,dxdy$$

which is simply the quadratic variation condition, also derived previously. The Euler equation is identical to that of the square Laplacian, namely the biharmonic equation $\nabla^4 f = 0$. The natural boundary conditions are different, however. They are given by

$$-\nabla^2 f + \left(f_{xx}x_s^2 + 2f_{xy}x_s y_s + f_{yy}y_s^2\right) = 0,$$
$$\frac{\partial}{\partial n}\nabla^2 f + \frac{\partial}{\partial s}\left(f_{xx}x_n x_s + f_{xy}(x_n y_s + x_s y_n) + f_{yy}y_n y_s\right) = 0.$$

In the simple case of a square boundary, oriented with respect to the coordinate axes, the boundary conditions reduce to:

$$f_{yy} = 0$$
$$2f_{xxy} + f_{yyy} = 0$$

along the boundary segments parallel to the x axis, and

$$f_{xx} = 0$$
$$2f_{yyx} + f_{xxx} = 0$$

along the boundary segments parallel to the y axis.

These boundary conditions can be straightforwardly simplified to:

$$f_{yy} = 0$$
$$f_{xxy} = 0$$

along the boundary segments parallel to the x axis, and

$$f_{xx} = 0$$
$$f_{yyx} = 0$$

along the boundary segments parallel to the y axis.

We have thus answered our question. For both the square Laplacian and the quadratic variation, the Euler equations are identical in the interior. The only difference to be noted in the extremal solutions to the two functionals will be observed at the boundaries of the surfaces. When we look at examples of solving these equations, this difference will become important.

There is a second manner in which the minimal solutions to the functionals will differ, in part related to the difference in boundary conditions of the two solutions. While the form of the minimal surface under either functional is roughly the same, except at the boundaries, this minimal surface will be uniquely determined only to within an element of the null space of the functional. This will be an important factor in determining which functional is best suited to our problem, since we would like the boundary conditions provided by the stereo data to completely determine a unique solution. The null spaces of the two functionals differ greatly, since the null space of the quadratic variation is the space of all linear functions, while the null space of the square Laplacian is the much larger space of all harmonic functions. We shall consider the effect of this difference later.

4.5 The Best Functional

Given that the set of possible functionals forms a vector space spanned by the square Laplacian and the quadratic variation, what criteria can be applied to determine the best functional? Since the Euler equation for both of these basis operators is the biharmonic equation $\nabla^4 f = 0$, the same will be true of any other operator in the vector space. This implies that aside from boundary conditions at the edge of the region being interpolated, the surfaces provided by any operator in this space will be basically the same.

This being the case, the only other characteristic that can distinguish between the possible functionals is the size of their respective null spaces. Let us denote the null space of the square Laplacian by $\mathcal{N}_1$ (the space of all harmonic functions) and the null space of the quadratic variation by $\mathcal{N}_2$ (the space of all linear functions). Note that $\mathcal{N}_2$ is a subspace of $\mathcal{N}_1$. Now the null space for any linear combination of these two operators must contain at least the space spanned by the intersection of the two null spaces $\mathcal{N}_1$ and $\mathcal{N}_2$. Hence, the null space of any other operator must consist at least of the linear functions. Thus, no possible operator can have a null space smaller than that corresponding to quadratic variation.

The importance of the null space is that it helps determine the family of surfaces that are minimal under the functional. The requirement we impose on the best functional is that the member of this family corresponding to the minimal surface be uniquely determined, when combined with the requirement that the surface must pass through the known points provided by the stereo algorithm. Clearly the smaller the size of the null space, the fewer the requirements we must impose on the output of the stereo algorithm in order to insure a unique solution.

We may view this criterion in the following manner. Initially, we start with the space of all possible functions, namely, the space of all second differentiable functions of two real variables, denoted $C^2(\Re^2)$. If we restrict our attention to those surfaces that pass through the boundary conditions imposed by the stereo or structure-from-motion data, we define a convex subset U of this space. If we define a functional on this space, the set of surfaces that are minimal under the functional are given by

$$\{v + w \mid \ w \in W\}$$

where

$$W = \{v - u \mid \ u \in U\} \cap \mathcal{N},$$

for some minimal surface $v \in U$. We are guaranteed a unique solution to the interpolation problem if W is empty (or equivalently, consists only of the null surface, defined to be zero everywhere). The key question becomes: can we have two surfaces that fit through the known points, have the same measure of surface consistency (the same value as measured by the functional) and differ by an element of the null space? If not, we are done, as the minimal surface is then guaranteed to be unique. Thus, the structure of the boundary conditions provided by the stereo algorithm (or the structure-from-motion algorithm) may be important in deciding which functional is more suitable. Clearly, the smaller the subspace of minimal surfaces, the more likely we are to have a unique minimal surface fitting the known data, as the set W is more likely to be empty.

Recall that the null space of the square Laplacian

$$\Theta(f) = \left\{ \int \int \left(\nabla^2 f \right)^2 dx dy \right\}^{\frac{1}{2}}$$

is the set of all harmonic functions. We wish to know what form of the boundary conditions will uniquely determine the harmonic function. This problem is known as the Dirichlet problem in

classical analysis, and it has long been known that if the boundary conditions consist of a series of closed, bounded Jordan curves, then the harmonic function is uniquely determined. These are, of course, sufficient, but not necessary conditions. It would appear, however, from these conditions that it is unlikely that the boundary conditions provided by the stereo algorithm will be sufficient to uniquely determine the component of the null space. This follows from the observation that the stereo algorithm is capable of providing boundary values at scattered points in the image, corresponding to the zero-crossings of the convolved image, while the Dirichlet problem is uniquely determined if the boundary values form a closed, bounded Jordan curve. Thus, in the case of the square Laplacian, the best we can do is determine a family of most consistent surfaces, which differ by harmonic functions. Referring back to our earlier question, we see that in this case, we could have two (or more) surfaces which fit through the known points, have the same measure of surface consistency, and differ by an element of the null space. The variation in such a family of surfaces is not consistent with our intuitive notion of indistinguishable surfaces, that is, the difference in the shape of two surfaces which have identical minimal values for the square Laplacian measured over the surface can be noticably large. As a consequence, we consider the square Laplacian to be a poor choice for the functional.

On the other hand, the null space of the quadratic variation

$$\Theta(f) = \left\{ \int \int \left(f_{xx}^2 + 2f_{xy}^2 + f_{yy}^2 \right) dx dy \right\}^{\frac{1}{2}}$$

is the set of all linear functions. The boundary conditions required in this case to uniquely determine the component of the null space are much simpler. In particular, if the stereo algorithm provides at least three non-colinear points, the element of the null space is uniquely determined to be the null surface (the surface which is zero everywhere). It is clear that in almost all imaging situations, the stereo algorithm will be capable of providing the necessary boundary conditions, and thus the most consistent surface is uniquely determined.

Thus, we have seen that the only possible functionals that can be used to measure the *surface consistency constraint* form a vector space spanned by the square Laplacian operator and the quadratic variation operator. The minimal surface for any such operator satisfies the biharmonic equation in the interior of the region being interpolated, but along the boundaries of the region it may satisfy different differential equations than the minimal solution of any other operator. In general, this implies that the solution surfaces corresponding to different operators will generally differ in shape

only near the boundaries. To distinguish between possible operators, we examined the form of their null spaces. We showed that the operator with the smallest null space was the quadratic variation. Further, the stereo data is in general sufficient to uniquely determine the component of the null space corresponding to the minimal surface. That is, the surface that minimizes the quadratic variation, subject to passing through the known points provided by the stereo or structure-from-motion algorithms, is uniquely determined.

4.6 The Computational Problem

By combining the results of the last two chapters, it is now possible to state the computational theory of the problem of interpolating visual surface information.

The Interpolation of Visual Information: Suppose one is given a representation consisting of surface information at the zero-crossings of a Primal Sketch description of a scene (this could be either from the Marr-Poggio stereo algorithm, or from the Ullman structure-from-motion algorithm, or both). Within the context of the visual information available, the best approximation to the original surface in the scene is given by the minimal solution to the quadratic variation in gradient (or surface orientation)

$$\Theta(f) = \left\{ \int \int \left(f_{xx}^2 + 2f_{xy}^2 + f_{yy}^2 \right) dx dy \right\}^{\frac{1}{2}}.$$

Such approximations are guaranteed to be uniquely "best" to within an element of the null space of the functional Θ. In the case of quadratic variation, the null space is the set of all linear functions. Provided the set of known points supplied by the stereo algorithm or by the structure from motion algorithm includes at least three non-colinear points, the component of the surface due to the null space is uniquely determined to be the null surface. Hence, the surface most consistent with the visual information is uniquely determined.

It is worth noting that although the above statement is phrased in terms of zero-crossings obtained from images convolved with $\nabla^2 G$ filters, the heart of the statement is much broader in scope. The key point is that to interpolate any surface representation which contains explicit information only at sparse points in the representation, we need to find the "most conservative" surface consistent with the input information. This implies that between the known surface points, the surface should vary as little as possible. Thus, whether those known points correspond to zero-crossings, edges, or

some other basic descriptor of image changes, the surface interpolation algorithm should construct the surface which minimizes variation in the surface between the known points.

It is interesting to compare the criteria for surface interpolation developed here, as well as the specific theory of surface interpolation stated above with the work of Barrow and Tenenbaum, [1981].

5. Constrained Optimization

Our goal throughout this paper has been to find the surface that best fits the known data provided by the stereo algorithm or the structure-from-motion algorithm. In the preceeding sections, we saw that such a "best" surface exists and is characterized as the surface that minimizes the functional of quadratic variation, measured over the surface. The problem we address now is how to find that minimal surface. What is meant by "finding the minimal surface"? Our goal is to derive an algorithm that computes explicit surface values (such as depth, or relative depth) at all points on a discrete grid, m points on a side. (That is, the scene will be partitioned into an $m \times m$ grid, and to each grid point, we want to associated a surface value.)

In general terms, we are seeking an algorithm to solve an optimization problem — we want to compute the values of a set of parameters that optimize some function. In our case, the parameters correspond to the surface values at the grid points, and the function to be optimized is the measure of a discrete correlate to quadratic variation over the surface. We will restrict our attention to the class of optimization algorithms that satisfy three simple criteria of biological feasibility — parallelism, local-support, and uniformity. These three criteria, together with the form of the input data — scattered contours of known points — preclude many of the possible techniques for solving an optimization problem, but also suggest the use of techniques such as those of mathematical programming.

5.1 The Role of Algorithmic Criteria

An essential problem for any computational theory about early visual processing is to determine the implicit assumptions used by the visual system to perform the computation. These are valid assumptions about the environment that are explicitly incorporated into the computation. Ullman's rigidity assumption in visual motion perception [Ullman, 1979a], Marr and Hildreth's condition of linear variation and spatial coincidence assumption [Marr and Hildreth, 1980], and Marr and Poggio's

assumptions of uniqueness and continuity [Marr and Poggio, 1979] are three examples. Such assumptions may be considered as computational constraints on the problem.

There is a second set of criteria that may be applied to any theory and, more importantly, to any algorithm for a theory. They deal with the requirement of biological feasibility, and are important if one is to describe a model of the human system. They will be termed *algorithmic criteria.* Ullman [1979b] has listed a number of such criteria that should apply to any biologically feasible algorithm. A similar set is briefly sketched here.

Parallelism

The need to process large amounts of input data in short amounts of time implies the use of computations that can be implemented in a parallel manner, using a large number of interconnected processors.

Local Support

If the number of processors involved in the computation is large, it becomes infeasible to connect each one to all of the others. Rather, there should only be local connections between the processors. Here, "local" means not only that the number of connections be small, but also that since the information being processed has a two-dimensional plane as an underlying coordinate system, the connections should also be local in a spatial sense. If the support of a function, defined on a two-dimensional grid, is the set of points on the grid that contribute in a non-trivial manner to the computation of the function, then our requirement is that the processors implementing our computation must have local support.

Uniformity

One final consideration, though not as critical as the first two, concerns the uniformity of the processors. If it is possible, an algorithm that utilizes parallel networks of identical processors will be favored over other algorithms. Such a requirement is not crucial, however.

Although the original motivation for such restrictions on an algorithm arises from consideration of the human visual system and restrictions of biological feasibility, they could apply equally well to other types of image processing systems. As such, they are taken as general criteria for the computations to be investigated, regardless of whether the algorithm serves as a model of the human system. In designing algorithms to solve a particular visual process, the first step is to seek a method that

solves the problem. Having done so, one can then consider its applicability in light of the criteria outlined above, and possible modifications to the algorithm in order to satisfy those criteria.

5.2 Mathematical Programming

The surface interpolation problem, as we have developed it, can be viewed as an optimization problem; that is, the solution to the surface interpolation problem is equivalent to the minimal point of an objective hypersurface. There is a large body of literature on methods for finding the solution to optimization problems in general. In considering which one to apply to our problem, we take two factors into account. The first is the form of the input data supplied by stereo (and possibly also structure-from-motion). The key point is that the set of known points will generally consist of a series of zero-crossing contours, along which the depth is known. These contours are not closed, since the horizontal components will have no disparity value, and hence no depth value, assigned to them. Further, they tend to be scattered at random rather than distributed uniformly over the grid. (This removes many methods from further consideration.) The second factor is the architecture of the possible algorithm, outlined by the algorithmic criteria of the previous section. As a consequence of these two factors, many of the possible methods, while perfectly valid solutions mathematically, are not readily applicable to our problem. A comprehensive review of the types of methods may be found in Schumaker [1976] (see also Grimson [1980, 1981b]).

Given that an algorithm used to solve the visual surface interpolation problem must be local, parallel and uniform, and must be capable of dealing with scattered input data, one of the best methods to use is *mathematical programming*, and in particular, nonlinear programming. Ullman [1979b] has shown that many problems of relaxation and constrained optimization can be solved by local nonlinear programming processes (see also Hummel and Zucker [1980]). Indeed, a method similar to that outlined here was used by Ullman in solving the motion correspondence problem [Ullman, 1979a].

Recall that the problem with which we are faced is to find the surface that minimizes a functional measuring surface consistency. The most likely candidate for this functional is the quadratic variation. The boundary conditions with which the surface must agree are depth values along the zero-crossings, given either by the Marr-Poggio stereo algorithm or the Ullman structure-from-motion algorithm. These boundary conditions can be met in one of two ways. If the surface is

required to fit exactly through the boundary points, the problem is one of surface interpolation. If the surface is required only to pass near the known points, while minimizing some error function, the problem is one of surface approximation. In the following sections, both problems will be examined.

Two cases of optimization will be examined: unconstrained optimization, which is applicable to the approximation problem, and constrained optimization, which is applicable to the interpolation problem. Standard algorithms for computing the solution to the optimization problem for each case are sketched below. For more details on mathematical programming, see for example Luenberger, [1973].

The Conjugate Gradient Algorithm

Starting at any point $\mathbf{x}_0 \in E^n$ define $\mathbf{d}_0 = -\mathbf{g}_0 = \mathbf{b} - \mathbf{Q}\mathbf{x}_0$ and

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$$

$$\alpha_k = -\frac{\mathbf{g}_k^T \mathbf{d}_k}{\mathbf{d}_k^T \mathbf{Q} \mathbf{d}_k}$$

$$\mathbf{d}_{k+1} = -\mathbf{g}_{k+1} + \beta_k \mathbf{d}_k$$

$$\beta_k = \frac{\mathbf{g}_{k+1}^T \mathbf{Q} \mathbf{d}_k}{\mathbf{d}_k^T \mathbf{Q} \mathbf{d}_k}$$

where $\mathbf{g}_k = \mathbf{Q}\mathbf{x}_k - \mathbf{b}$. ∎

We will apply this algorithm, for the case of unconstrained optimization, to the problem of visual surface interpolation in the next section. Because the algorithm is solving an unconstrained optimization problem, it will be applicable to the surface approximation problem, where the surface is required to pass near, but not necessarily through, the known points.

Gradient Projection Algorithm

The algorithm may be summarized as follows.

1. Find the subspace of active constraints M, and form the matrix $\mathbf{A}_p$.
2. Calculate the projection matrix $\mathbf{P}_k = \left[\mathbf{I} - \mathbf{A}_p^T\left(\mathbf{A}_p\mathbf{A}_p^T\right)^{-1}\mathbf{A}_p\right]$ and the direction vector $\mathbf{d} = -\mathbf{P}_k \nabla f(\mathbf{x})^T$.
3. If $\mathbf{d} \neq \mathbf{0}$, find the scalar c_1 that maximizes

$$\{\alpha\colon\ \mathbf{x} + \alpha\mathbf{d} \text{ is feasible}\}$$

and the scalar c_2 that minimizes

$$\{f(\mathbf{x} + \alpha\mathbf{d})\colon\ 0 \leq \alpha \leq c_1\}$$

as a function of α. Set $\mathbf{x}$ to $\mathbf{x} + c_2\mathbf{d}$ and return to (1).

4. If $\mathbf{d} = \mathbf{0}$, find $\beta = -\left(\mathrm{A}_p\mathrm{A}_p^T\right)^{-1}\mathrm{A}_p\nabla f(\mathbf{x})^T$. If $\beta_j \geq 0$, for all j corresponding to active inequalities, stop, as $\mathbf{x}$ satisfies the Kuhn-Tucker conditions. Otherwise, delete the row from A_p corresponding to the inequality with the most negative component of β and return to (2). ∎

We will apply this algorithm, for the case of constrained optimization, to the problem of visual surface interpolation in the next section. Because the algorithm is solving a constrained optimization problem, it will be applicable to the surface interpolation problem, where the surface is required to pass through the known points.

6. The Interpolation Algorithm

The algorithms of the previous section can now be applied to the problem at hand, the interpolation (or approximation) of visual surfaces from the stereo data. Recall that the interpolation problem was stated as:

The Interpolation of Visual Information: Suppose one is given a representation consisting of surface information at the zero-crossings of a Primal Sketch description of a scene (this could be either from the Marr-Poggio stereo algorithm, or from the Ullman structure-from-motion algorithm, or both). Within the context of the visual information available, the best approximation to the original surface in the scene is given by the minimal solution to the quadratic variation in gradient

$$\Theta(s) = \left\{\int\int \left(s_{xx}^2 + 2s_{xy}^2 + s_{yy}^2\right) dxdy\right\}^{\frac{1}{2}},$$

(where s denotes a surface). Such approximations are guaranteed to be uniquely "best" to within an element of the null space of the functional Θ. In the case of quadratic variation, the null space is the set of all linear functions. Provided the set of known points supplied by the stereo algorithm or by the structure-from-motion algorithm includes at least three non-colinear points, the component of the surface due to the null space is uniquely determined to be the null surface. Hence, the surface most consistent with the visual information is uniquely determined.

We shall consider solving this optimization problem both in the case of interpolation (the surface passes exactly through the data) and in the case of approximation (the surface passes near the data). Although the algorithms could be either applied to the square Laplacian or to the quadratic variation, we shall examine only the case of the quadratic variation.

6.1 Conversion to the Image Domain

The problem, as stated, lies clearly within the domain of continuous functions. Yet this is not appropriate to the case at hand. In order to establish an algorithm for transforming the visual information into a representation of the surface shape, a number of conversions must take place.

The first point to note is that the functional $\Theta(s)$ consists of a square root. (Note that we will use s to denote the surface which we are fitting to the known points, to distinguish it from the notation of s used in the previous chapter to denote the objective function.) However, clearly any function which minimizes the functional $\Theta(s)$ also minimizes the functional $\Theta^2(s)$, and vice versa, provided that the functional is always positive in value. Hence, without loss of generality, one may consider the minimization of

$$\Theta(s) = \int\int \left(s_{xx}^2 + 2s_{xy}^2 + s_{yy}^2\right) dxdy.$$

Throughout this section, this will be the functional to be minimized.

In order to determine the structure of the algorithm, one must address the issue of the form of the output representation, since that will have a major effect on the actual algorithm. In this case, it is desired that the surface information be specified only at particular places within the representation of the scene. This will be accomplished by requiring that the interpolation algorithm compute explicit depth values at all locations within a Cartesian grid of uniform spacing. Although both the spatial resolution of the grid and the resolution of the depth information stored within that grid should be determined, it is considered that such parameters are not critical to the development of the algorithm. Hence, these parameters will be assigned arbitrary values.

The continuous functional must now be converted to a form applicable to a discrete grid. Without loss of generality, assume that the grid is of size $m \times m$. Each point on the grid may be represented by its coordinate location, so that the point (i, j) corresponds to the grid point lying on the i^{th} row and the j^{th} column. At each point (i, j) on the grid, a surface value may be represented by $s_{(i,j)}$. Each such surface value may be considered as an independent variable, subject to the constraints of the problem, of course. Using either row major order or column major order, these variables $s_{(i,j)}$ may be considered as a vector of variables, denoted $\mathbf{s} = \{s_{(0,0)}, s_{(0,1)}, \cdots, s_{((m-1),(m-1))}\}$. (For clarity, a straightforward transformation from the doubly-indexed grid coordinates into a singly-indexed vector coordinate can be established. For example, the grid point (i, j) can be mapped to the vector point $k = mi + j$, and the vector point k can be mapped to the grid point (i, j) —

$(\lfloor k/m \rfloor, k - m\lfloor k/m \rfloor)$.) It is this vector which will be modified using the non-linear programming algorithms, and the final value of which will form the solution to the optimization problem and thus correspond to the desired interpolated surface.

Having converted the surface function to a discrete grid format, it is now necessary to convert the objective function of the optimization problem to a discrete format. This means that the differential operators must be converted to difference operators. There are many possible discrete approximations to the differential operators. We choose to use the following approximations [Abramowitz and Stegun, 1965, p. 884].

The second partial derivative in the x direction may be approximated by

$$\frac{\partial^2 s_{(i,j)}}{\partial x^2} = \frac{1}{h^2}\left[s_{(i+1,j)} - 2s_{(i,j)} + s_{(i-1,j)}\right] + O(h^2)$$

where h is the grid spacing, and $O(h^2)$ indicates that the approximation is valid to terms of order h^2. Similarly, the second partial derivative in the y direction may be approximated by

$$\frac{\partial^2 s_{(i,j)}}{\partial y^2} = \frac{1}{h^2}\left[s_{(i,j+1)} - 2s_{(i,j)} + s_{(i,j-1)}\right] + O(h^2).$$

The cross second partial derivative can be approximated by

$$\frac{\partial^2 s_{(i,j)}}{\partial x \partial y} = \frac{1}{4h^2}\left[s_{(i+1,j+1)} - s_{(i+1,j-1)} - s_{(i+1,j-1)} + s_{(i-1,j-1)}\right] + O(h^2).$$

Note that such approximations have frequently been used in the image processing literature, (for example, see the reviews of Davis [1975], Rosenfeld and Kak [1976], Pratt [1978]). Little is known of the affect of these approximations on the behavior of the result.

Having converted the surface function and the differential operators, one must convert the double integral to a discrete equivalent. This can easily be done, by using a double summation over the finite difference operators applied to the discrete grid. One minor point is noted. While it is straightforward to form the discrete equivalent to the double integrals $\int\int s_{xx}^2 dxdy$ and $\int\int s_{yy}^2 dxdy$, the cross term $2\int\int s_{xy}^2 dxdy$ is handled differently. In particular, consider a second grid, superimposed on the first, which has twice the spatial resolution of the first (that is, all integer points are represented as are all points $(\frac{i}{2}, \frac{j}{2})$). For the cross term, we shall apply the finite difference operator to all half integral points on this finer grid. The combination of these operators yields the

discrete objective function:

$$\text{minimize} \quad \sum_{i=1}^{m-2}\sum_{j=0}^{m-1}\left(s_{(i-1,j)} - 2s_{(i,j)} + s_{(i+1,j)}\right)^2 + \sum_{i=0}^{m-1}\sum_{j=1}^{m-2}\left(s_{(i,j-1)} - 2s_{(i,j)} + s_{(i,j+1)}\right)^2 + 2\sum_{i=0}^{m-2}\sum_{j=0}^{m-2}\left(s_{(i,j)} - s_{(i+1,j)} - s_{(i,j+1)} + s_{(i+1,j+1)}\right)^2.$$

Finally, the characterization of the constraints must be considered. The case of interpolation will be considered first, where the interpolated surface is required to pass through the known points. Let $\mathcal{S} = \{(i,j) \mid$ there is a known depth value at the grid point $(i,j)\}$ be the set of grid points for which a depth value is known. Then the constraints on the optimization problem have the form $s_{(i,j)} - c_{(i,j)} = 0$ for all points (i,j) in the set $\mathcal{S}$, and where the $c_{(i,j)}$'s are a set of constants reflecting the stereo data. Note that the set of constraints are all equality constraints.

6.2 The Gradient Projection Interpolation Algorithm

It is now possible to consider applying the gradient projection method to this problem:

$$\text{minimize} \quad \sum_{i=1}^{m-2}\sum_{j=0}^{m-1}\left(s_{(i-1,j)} - 2s_{(i,j)} + s_{(i+1,j)}\right)^2 + \sum_{i=0}^{m-1}\sum_{j=1}^{m-2}\left(s_{(i,j-1)} - 2s_{(i,j)} + s_{(i,j+1)}\right)^2 + 2\sum_{i=0}^{m-2}\sum_{j=0}^{m-2}\left(s_{(i,j)} - s_{(i+1,j)} - s_{(i,j+1)} + s_{(i+1,j+1)}\right)^2.$$

$$\text{subject to} \quad s_{(i,j)} - c_{(i,j)} = 0, \ \forall (i,j) \in \mathcal{S}.$$

To apply the method of gradient projection, it is necessary to determine the set of active constraints, and the projection matrix onto the subspace spanned by the active constraints. Clearly, since all the constraints are equality constraints, they are all active at every iteration. Thus, the matrix A^p

(where $p = |\mathcal{S}|$) has rows consisting of a 1 in the position corresponding to the grid point (i, j) for $(i, j) \in \mathcal{S}$ and 0's elsewhere. One can easily show that $\mathbf{A}_p\mathbf{A}_p^T = \mathbf{I}$ and that $\mathbf{A}_p^T\mathbf{A}_p = \delta_{\mathcal{S}}$ where $\delta_{\mathcal{S}}$ is a matrix consisting on 0's except for those rows corresponding to a point in $\mathcal{S}$, such rows containing a 1 for the diagonal element. Thus, the projection matrix $\mathbf{P} = \mathbf{I} - \delta_{\mathcal{S}}$ consists of all 0's except for the diagonal elements in those rows corresponding to a grid point not in $\mathcal{S}$, such elements being 1. The effect of the projection matrix $\mathbf{P}$ is to ignore any components of the direction vector $\mathbf{d}$ corresponding to a known point, while preserving all other components, unaltered.

Recall that the direction vector $\mathbf{d}$ is determined by the projection of the negative gradient of the objective function. By expanding the double summation and performing the differentiation, the components of the gradient of the objective function are given, in this case, by the following:

For all elements in the center of the grid, apply the following stencil to the grid representation of the surface function $\mathbf{s}$:

$$\begin{bmatrix} & & 2 & & \\ & 4 & -16 & 4 & \\ 2 & -16 & 40 & -16 & 2 \\ & 4 & -16 & 4 & \\ & & 2 & & \end{bmatrix}.$$

By this, we mean that given a two-dimensional grid representation of the current surface approximation, s, the value of the component of the gradient of the objective surface at some point (i, j) on the grid is obtained by applying the above stencil centered over that point (i, j), multiplying the value of each of the stencil points with the value of the surface at that point and summing these products. The value of the components of the gradient can be computed in this manner by applying the stencil to all points in the center of the grid.

Along the outer edges of the grid, the above stencil does not apply. Instead, a careful expansion of the gradient of the objective function shows that the following stencils should be used.

For elements in the corners of the grid, apply the following stencil (or its appropriate rotations and reflections) to the grid representation of the surface function $\mathbf{s}$:

$$\begin{bmatrix} 2 & & \\ -8 & 4 & \\ 8 & -8 & 2 \end{bmatrix}.$$

For elements along an outside row of the grid, one point removed from the corner, apply the following stencil (or its appropriate rotations and reflections) to the grid representation of the surface function **s**:

$$\begin{bmatrix} & 2 & \\ 4 & -12 & 4 \\ -8 & 20 & -12 \quad 2 \end{bmatrix}.$$

For elements along an outside row of the grid, more than one point removed from any corner, apply the following stencil (or its appropriate rotations and reflections) to the grid representation of the surface function **s**:

$$\begin{bmatrix} & 2 & \\ 4 & -12 & 4 \\ 2 & -12 & 22 \quad -12 \quad 2 \end{bmatrix}.$$

For elements along a row second from the outside of the grid, located one element from each of two outside rows, apply the following stencil (or its appropriate rotations and reflections) to the grid representation of the surface function **s**:

$$\begin{bmatrix} & 2 & \\ 4 & -16 & 4 \\ -12 & 36 & -16 \quad 2 \\ 4 & -12 & 4 \end{bmatrix}.$$

For all other elements along a row second from the outside of the grid, apply the following stencil (or its appropriate rotations and reflections) to the grid representation of the surface function **s**:

$$\begin{bmatrix} & 2 & \\ 4 & -16 & 4 \\ 2 & -16 & 38 \quad -16 \quad 2 \\ 4 & -12 & 4 \end{bmatrix}.$$

Thus, the direction vector has zero valued components at all points corresponding to known depth values, and non-zero valued components elsewhere, with value given by the result of convolving the above stencils with the current surface approximation. It is interesting to note that the stencil used in the interior of the grid is a finite difference approximation to the biharmonic equation $\nabla^4 s = 0$ [Abramowitz and Stegun, 1965, p.885]. This should not be surprising, since the Euler equation,

derived previously from the calculus of variations, was precisely this equation. Thus, we see that the quadratic programming algorithms implicitly solve the Euler equation.

We have determined the form of the direction vector, which specifies the direction in which to move in order to reduce the objective function and refine the surface approximation. To determine the amount to move in this direction, it is necessary to determine the minimum value of the objective function along this direction, that is, to determine the value of α such that

$$\sum_{i=1}^{m-2}\sum_{j=0}^{m-1}\Big(s_{(i-1,j)} - 2s_{(i,j)} + s_{(i+1,j)} + \alpha d_{(i-1,j)} - 2\alpha d_{(i,j)} + \alpha d_{(i+1,j)}\Big)^2$$
$$+\sum_{i=0}^{m-1}\sum_{j=1}^{m-2}\Big(s_{(i,j-1)} - 2s_{(i,j)} + s_{(i,j+1)} + \alpha d_{(i,j-1)} - 2\alpha d_{(i,j)} + \alpha d_{(i,j+1)}\Big)^2$$
$$+2\sum_{i=0}^{m-2}\sum_{j=0}^{m-2}\Big(s_{(i,j)} - s_{(i+1,j)} - s_{(i,j+1)} + s_{(i+1,j+1)} + \alpha d_{(i,j)} - \alpha d_{(i+1,j)} - \alpha d_{(i,j+1)} + \alpha d_{(i+1,j+1)}\Big)^2$$

is minimized. A straightforward application of calculus determines that this value for α is given by the ratio of $\alpha = \frac{a_1}{a_2}$ where

$$\alpha_1 = \sum_{i=1}^{m-2}\sum_{j=0}^{m-1}\left(s_{(i-1,j)} - 2s_{(i,j)} + s_{(i+1,j)}\right)\left(d_{(i-1,j)} - 2d_{(i,j)} + d_{(i+1,j)}\right)$$
$$+ \sum_{i=0}^{m-1}\sum_{j=1}^{m-2}\left(s_{(i,j-1)} - 2s_{(i,j)} + s_{(i,j+1)}\right)\left(d_{(i,j-1)} - 2d_{(i,j)} + d_{(i,j+1)}\right)$$
$$+2\sum_{i=0}^{m-2}\sum_{j=0}^{m-2}\left(s_{(i,j)} - s_{(i+1,j)} - s_{(i,j+1)} + s_{(i+1,j+1)}\right)\left(d_{(i,j)} - d_{(i+1,j)} - d_{(i,j+1)} + d_{(i+1,j+1)}\right)$$

and

$$\alpha_2 = \sum_{i=1}^{m-2}\sum_{j=0}^{m-1}\left(d_{(i-1,j)} - 2d_{(i,j)} + d_{(i+1,j)}\right)^2$$
$$+ \sum_{i=0}^{m-1}\sum_{j=1}^{m-2}\left(d_{(i,j-1)} - 2d_{(i,j)} + d_{(i,j+1)}\right)^2$$
$$+2\sum_{i=0}^{m-2}\sum_{j=0}^{m-2}\left(d_{(i,j)} - d_{(i+1,j)} - d_{(i,j+1)} + d_{(i+1,j+1)}\right)^2.$$

Thus, the algorithm is completely determined. The steps consist of:

0. Determine a feasible initial surface approximation (any surface approximation which contains the known stereo depth values $c_{(i,j)}$ at the known points $(i, j) \in \mathcal{I}$ will suffice).

1. Compute the gradient of the objective function by convolving the grid representation of the current surface approximation with the stencils listed above. Compute the direction vector by taking the negative of the gradient, setting any components corresponding to known depth points to zero.

2. Compute the scalar α which specifies the amount to move along the direction vector on the hypersurface defined by the objective function, by the formula given above.

3. Refine the surface approximation by incrementing the current surface approximation with the scaled direction vector.

4. Return to step (1) and continue until the magnitudes of all components of the direction vector are smaller than some constant ϵ.

6.3 Examples of Interpolation

We can demonstrate the effectiveness of the surface interpolation algorithm by considering the performance of the gradient projection algorithm on several examples. Although the previous discussion dealt specifically with applying the gradient projection algorithm to the quadratic variation, a similar analysis can be performed for other functionals such as the square Laplacian. (Recall that any feasible functional was a linear combination of these two functionals.)

To demonstrate both the effectiveness of the interpolation algorithm and the difference between the quadratic variation and the square Laplacian, we consider three synthetic examples in Figures 4-6. In Figure 4, the interpolation algorithm is given as boundary conditions a set of closed contours from a cylinder, oriented parallel to the x-axis. It can be seen that the surfaces obtained by minimizing the square Laplacian and the quadratic variation differ markedly along the edge of the region. This is to be expected for two reasons. In Section 5, we derived the Euler equations for the interpolation problem, a set of differential equations which must be satisfied by the minimal surface. The Euler equations for the interior of a region were identical for both the square Laplacian and the quadratic variation, namely the biharmonic equation. Along the edges of the region, however, the natural boundary conditions imposed different equations on the solution surface. This fact is reflected in Figure 4. The second reason for the different surfaces arises from the form of the stencils obtained in Section 6.2. The stencils to be applied at the edges of a region in the case of quadratic variation are numerically more stable than those to be applied in the case of the square Laplacian. (This may, in fact, simply be a reflection of the difference in Euler equations.) In either case, it can be seen from Figure 4 that while minimizing the quadratic variation results in a reasonable approximation to a cylinder, minimizing the square Laplacian yields less acceptable results.

In Figure 5, we illustrate a second synthetic example. In this case, the boundary conditions are points lying on a hyperbolic paraboloid, choosen at random so that the known points do not form closed contours. As in the previous case, it can be seen that while the surfaces obtained by minimizing the two functionals are very similar in the interior of the region, the surfaces differ drastically along

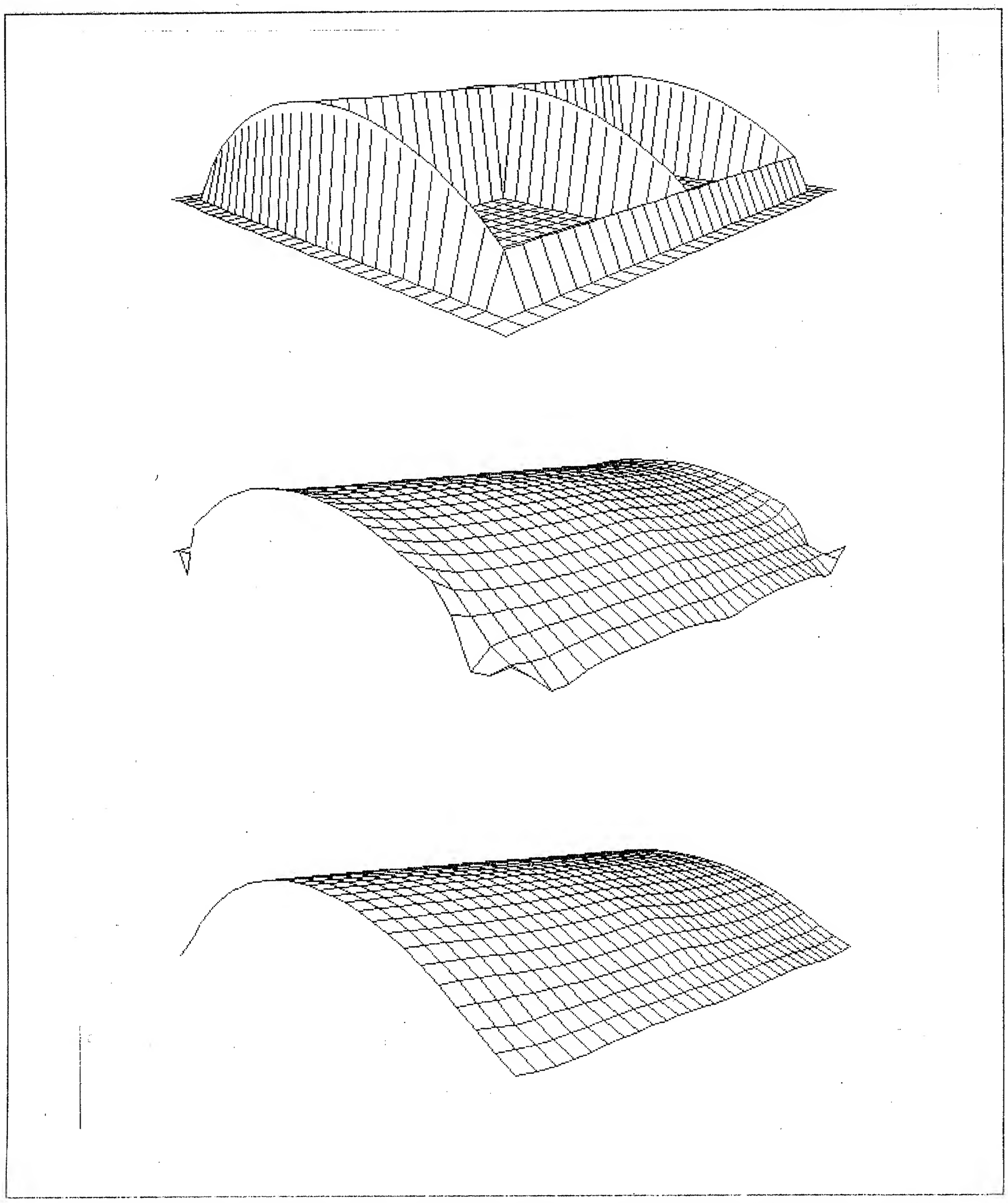

Figure 4. Synthetic Example. The top figure shows a synthetic set of boundary conditions, consistent with a cylinder aligned with the axes of the grid. The middle figure shows the surface obtained by applying the gradient projection algorithm to the square Laplacian functional. The bottom figure shows the surface obtained by applying the algorithm to the quadratic variation.

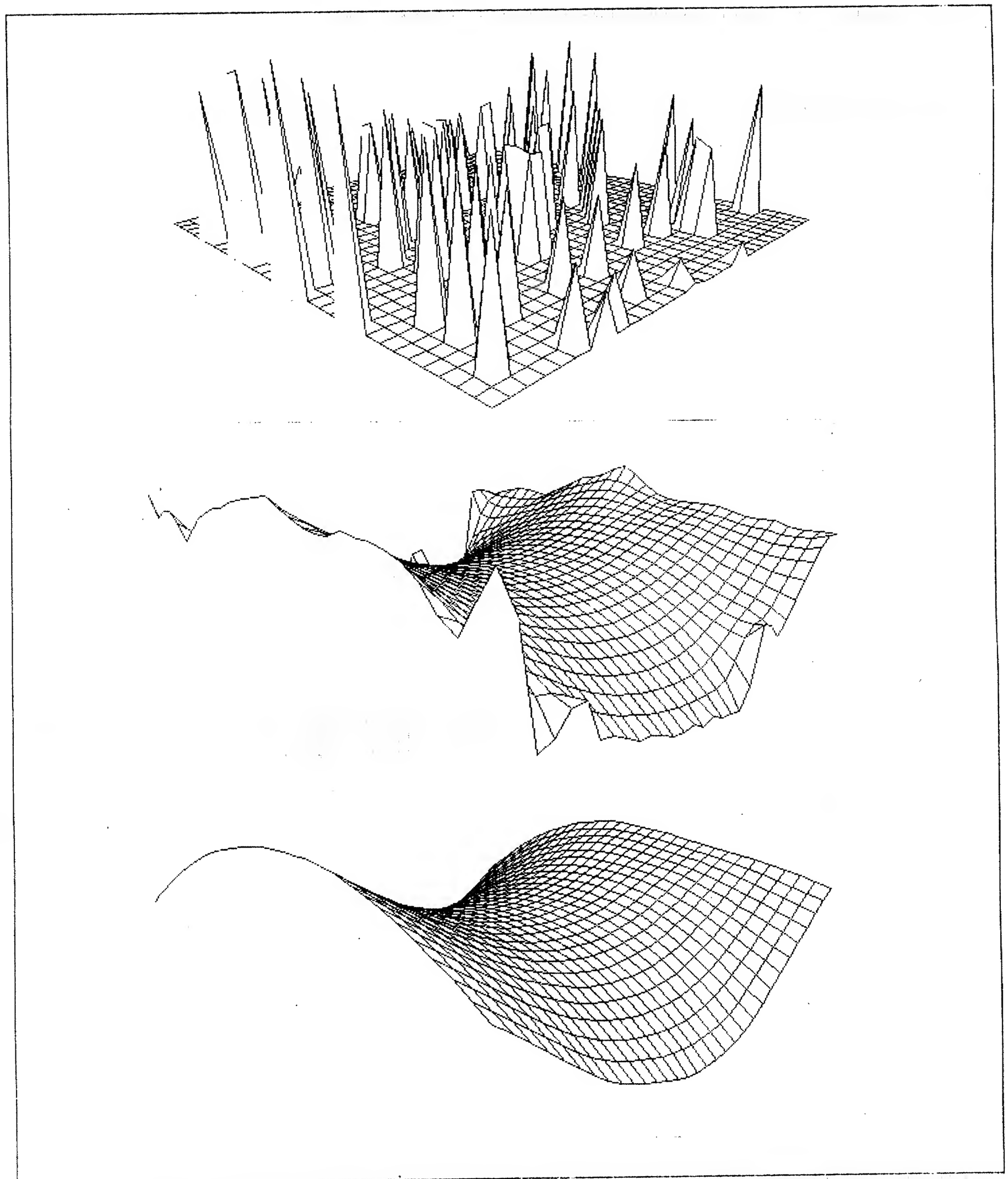

Figure 5. Synthetic Example. The top figure shows a synthetic set of boundary conditions, consistent with a hyperbolic paraboloid. The points are chosen at random with a density of 10 percent. The middle figure shows the surface obtained by applying the gradient projection algorithm to the square Laplacian functional. The bottom figure shows the surface obtained by applying the algorithm to the quadratic variation.

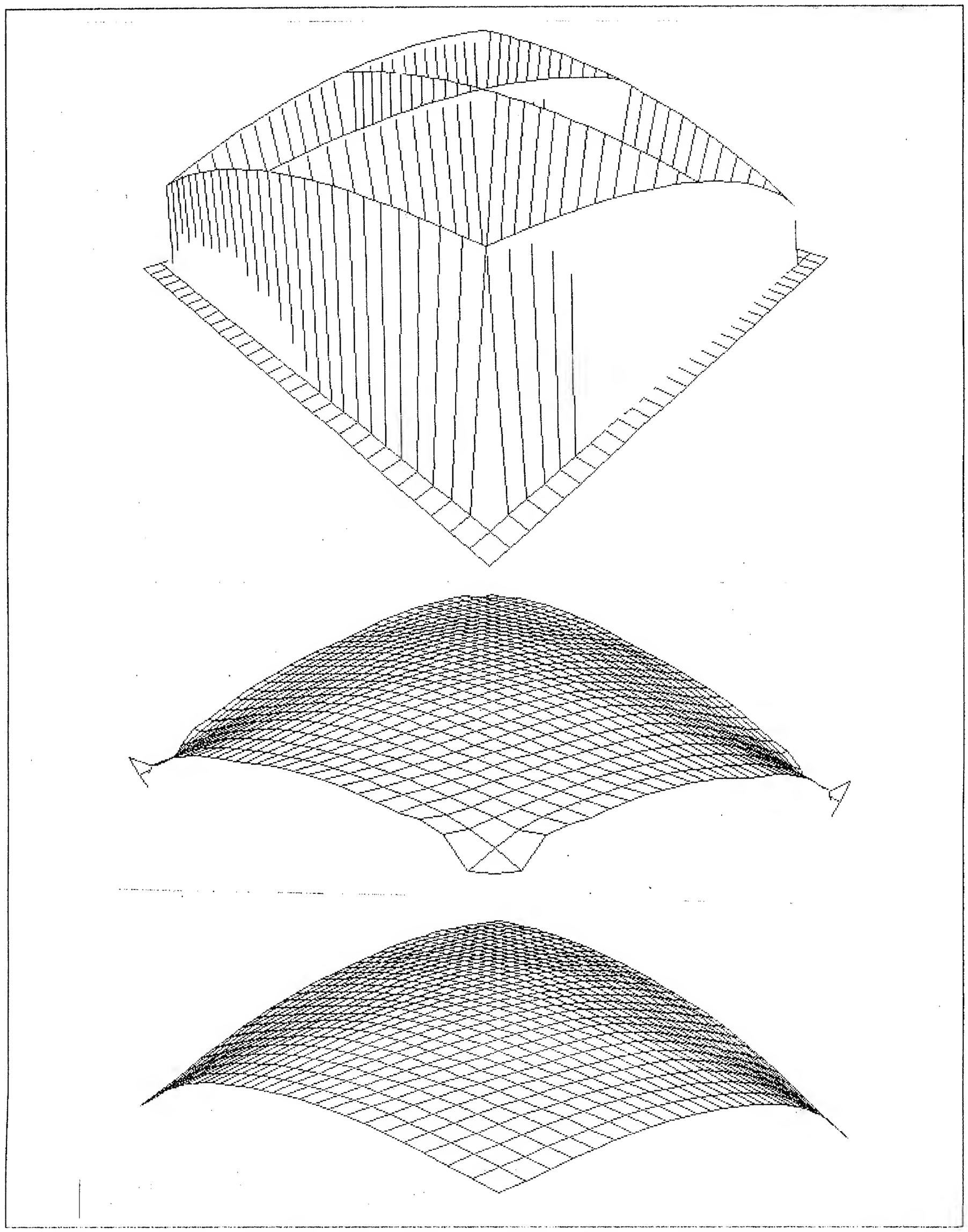

Figure 6. Synthetic Example. The top figure shows a synthetic set of boundary conditions, consistent with a cylinder not aligned with the axes of the grid. The middle figure shows the surface obtained by applying the gradient projection algorithm to the square Laplacian functional. The bottom figure shows the surface obtained by applying the algorithm to the quadratic variation.

the edges of the region. Again, minimizing the quadratic variation yields a reasonable approximation to a hyperbolic paraboloid.

In Figure 6, we illustrate a third synthetic example. In this case, the boundary conditions are again taken from a cylinder, here oriented at 45 degrees to the x-axis. As in the previous cylinder example, the major difference between the two surfaces occurs along the borders of the region and the minimization of quadratic variation yields a good approximation to the cylinder.

The interpolation algorithm was developed to account for the creation of complete surface descriptions from the sparse surface information provided by visual modules such as stereo. We can demonstrate the effectiveness of the interpolation theory by applying the algorithm to different stereo examples. In Figures 7-10, we illustrate the results of applying the surface interpolation algorithm to the output of the Grimson implementation [Grimson, 1981a, 1981b] of the Marr-Poggio stereo theory [Marr and Poggio, 1979] applied to a pair of stereo images. It should be noted that in these examples the interpolation algorithm was applied directly to the disparity values obtained by the stereo algorithm, without converting them to depth information. As a consequence, the displayed surfaces in the figures will not exactly reflect the shape of the surface, since an additional nonlinear transformation from disparity to depth is still required. For the purposes of illustrating the interpolation algorithm, however, the use of interpolated disparity values suffices, since the interpolation algorithm will preserve the general shape of the surfaces (that is, the sign of the surface curvature) as well as the relative differences in depth between different surfaces.

Figure 7 shows four stereo pairs of images, on which the algorithm was tested. Figure 8 shows the surface obtained for a wedding cake random dot stereogram. The four planar surfaces are clearly visible, although the effect of a small number of incorrect disparity values at the junctions of adjacent planes can be seen. Figure 9 shows the surface obtained for a spiral staircase random dot stereogram. Again, while the general shape of the spiral staircase is clearly apparent, the effect of a small number of incorrect disparity values can be seen. Figure 10 shows the surface obtained for the natural image of a coffee jar. As in the previous cases, the general shape of the surfaces are clearly evident. Not only is the jar sharply separated in disparity from the background plane (which is slightly slanted), but the overall shape of the jar can be distinguished. Figure 11 shows the surface obtained for the natural image of the Moore sculpture. As in the case of Figure 10, the general shape of the surface can be

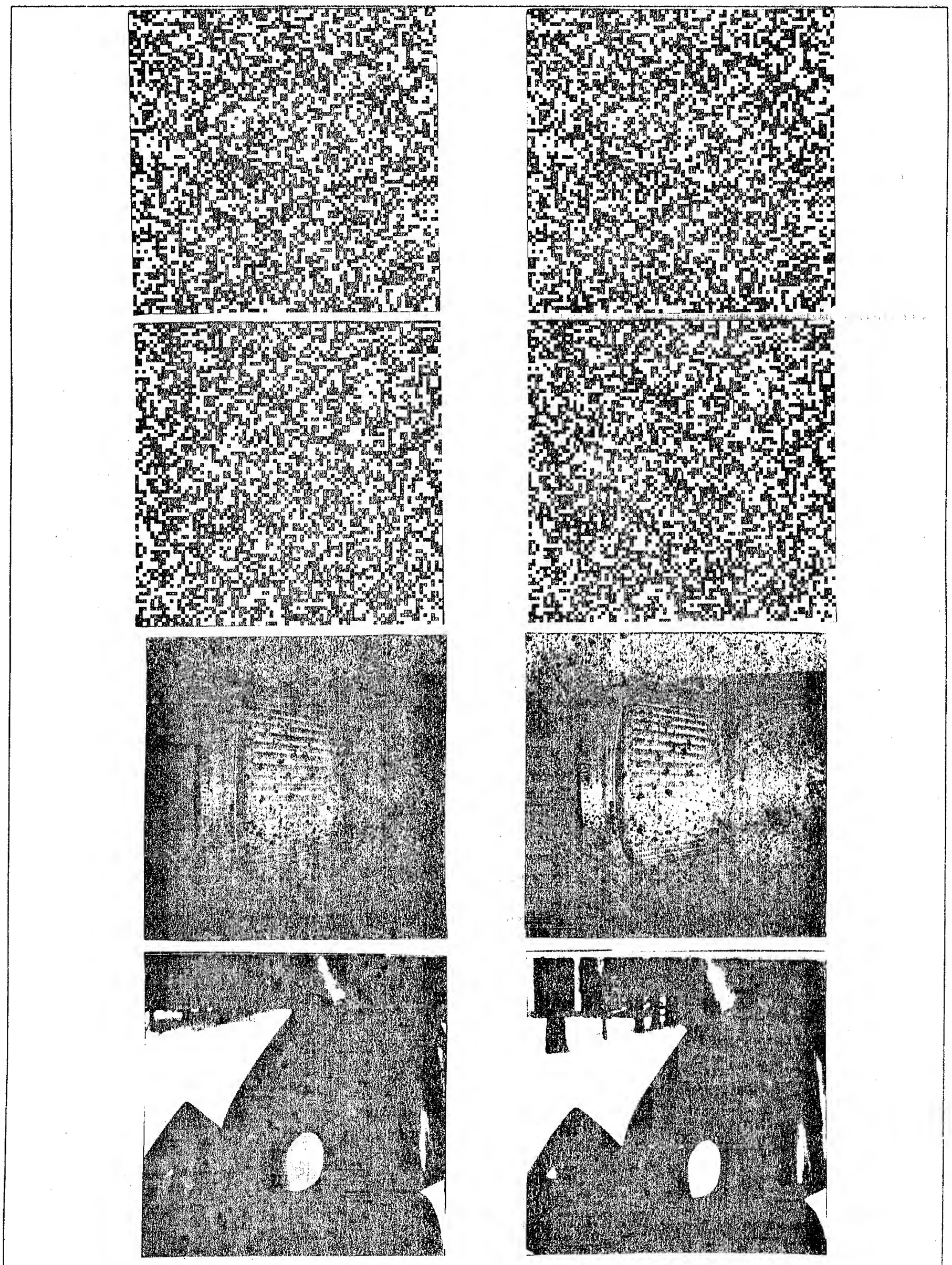

Figure 7. Examples of Stereo Images. The figures, from top to bottom, show the random dot stereogram of a wedding cake, the random dot stereogram of a spiral staircase, a natural image of a coffee bottle, and a natural image of a sculpture by Henry Moore.

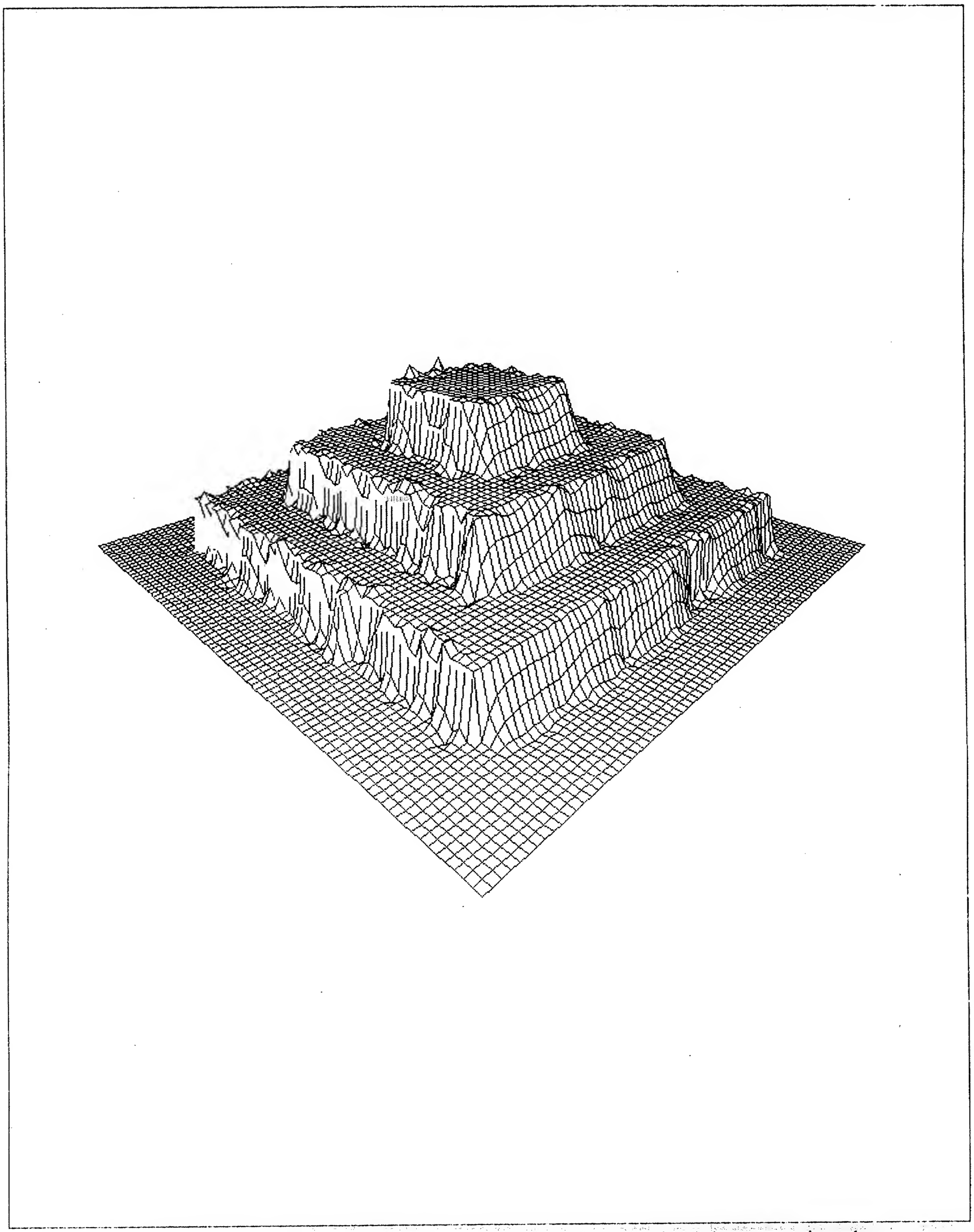

Figure 8. The Wedding Cake. The figure shows the surface obtained by processing the stereo pair with the Grimson implementation of the Marr-Poggio stereo algorithm, and interpolating the result using the quadratic variation.

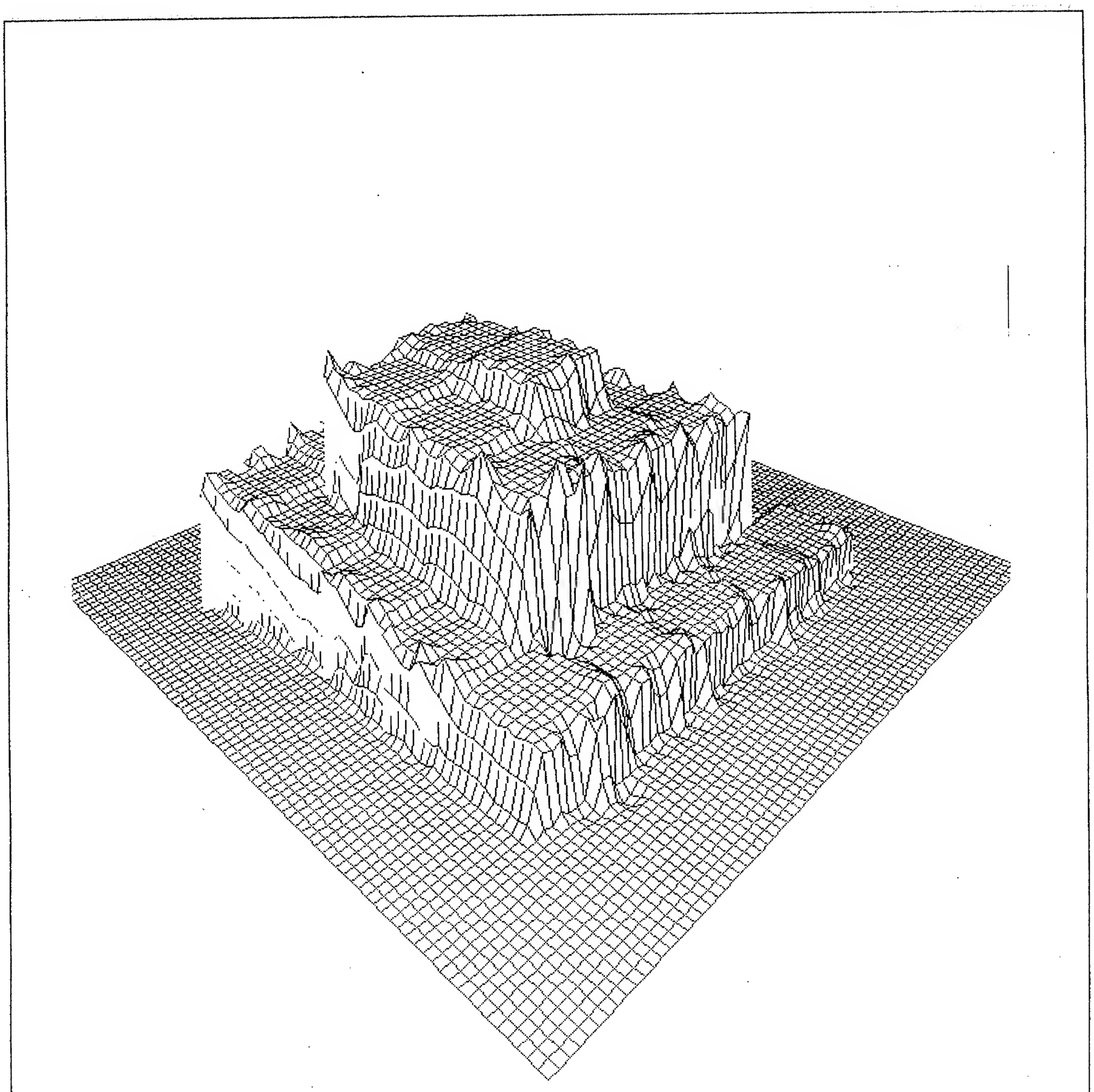

Figure 9. The Spiral Staircase. The figure shows the surface obtained by processing the stereo pair with the Grimson implementation of the Marr-Poggio stereo algorithm, and interpolating the result using the quadratic variation.

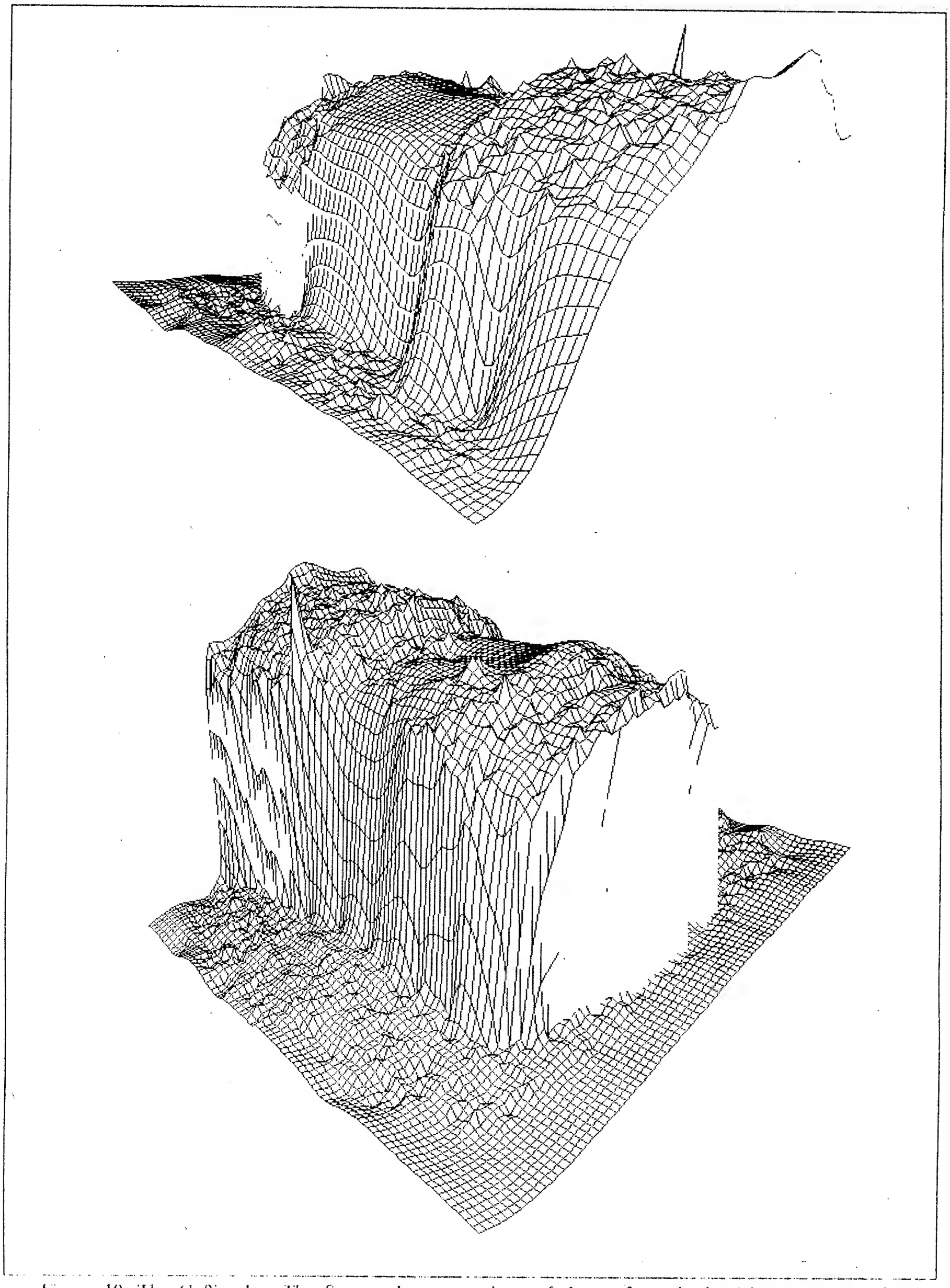

Figure 10. The Coffee Jar. The figures show two views of the surface obtained by processing the stereo pair of Figure 7 with the Grimson implementation of the Marr-Poggio stereo algorithm, and interpolating the result using the quadratic variation.

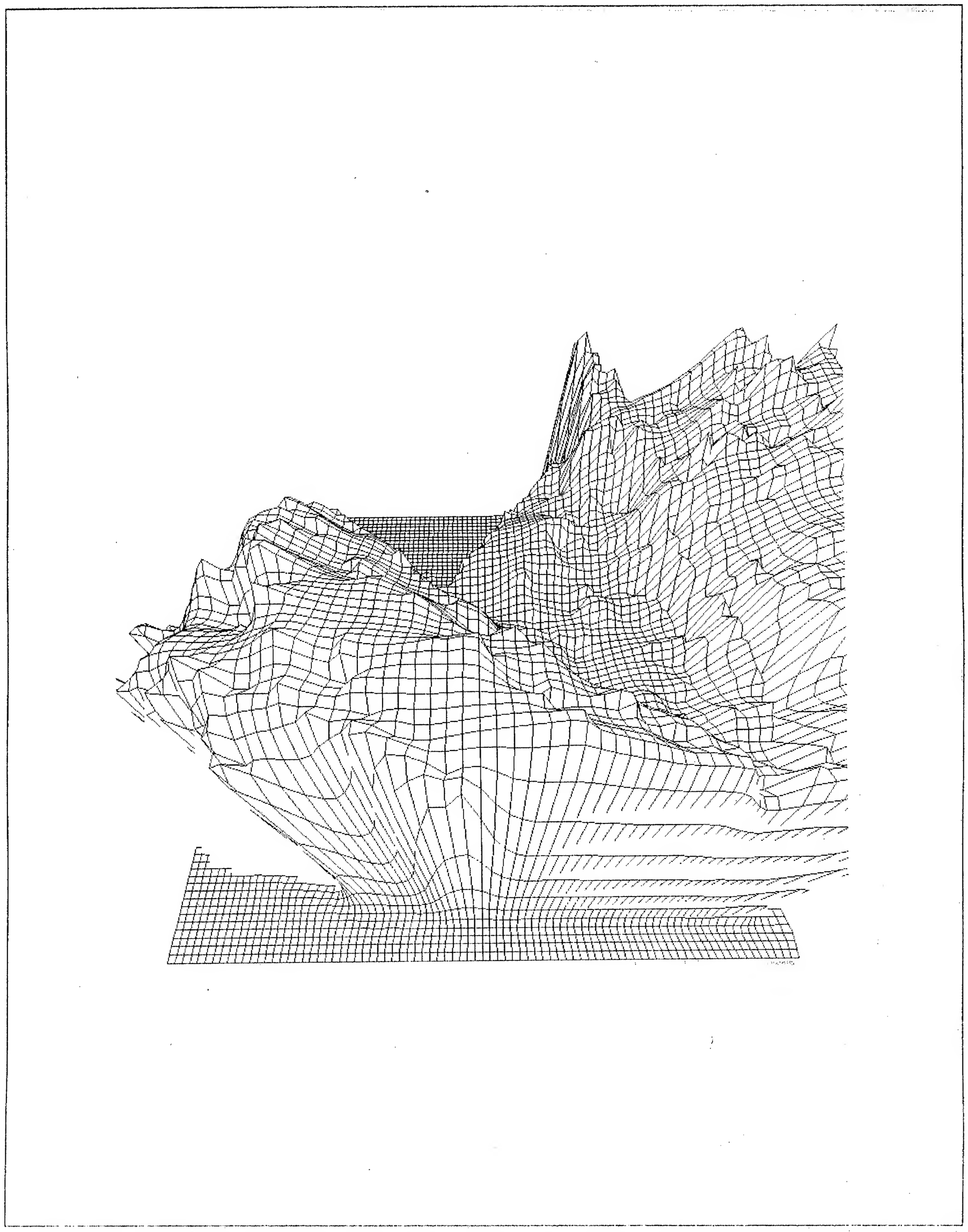

Figure 11. The Moore Sculpture. The figure shows a view of the surface obtained by processing the stereo pair of Figure 7 with the Grimson implementation of the Marr-Poggio stereo algorithm, and interpolating the result using the quadratic variation.

distinguished. Note that because no disparity values can be obtained for the hole in the center of the sculpture, the interpolation algorithm interpolates across the hole as if it were a uniform surface.

6.4 The Conjugate Gradient Approximation Algorithm

Previous sections have addressed the case of interpolating a surface through the known stereo depth values. In this section, the case of approximating a surface relative to the known stereo depth values is considered. There are several reasons for considering an approximation of the known depth values, rather than an exact fit through them. The first reason is that the accuracy of the stereo data may not be sufficient for the purpose of surface approximation. In particular, the algorithm outlined for performing the stereo computation yields disparity matches with an accuracy of one picture element. One must consider if such accuracy is sufficient. As well, one must consider the accuracy with which the zero-crossing positions reflect the location of a point of interest on an object in the scene. Since the operators which extract the zero-crossings have a non-infinitesimal spatial extent, it is possible that the zero-crossing positions undergo slight fluctuations in position, such fluctuations causing a small error in the disparity matches assigned by the algorithm. The second reason is that the stereo algorithm does occasionally make an incorrect match If an exact surface interpolation is required, such points will incorrectly cause a change in the shape of the surface, and the effect of such points can spread over a noticeable region of the surface reconstruction. By requiring a surface approximation, the effect of such "bad" disparity points can be minimized.

The basic notion is to combine a measure of "nearness of fit to the known points" with a measure of the consistency of the surface with the zero-crossing information. This can be accomplished by considering a penalty method unconstrained optimization problem. Here, the objective function to minimize is

$$\Theta(s) = \int\int \left(s_{xx}^2 + s_{yy}^2 + 2s_{xy}^2 \right) dxdy + \beta \sum_{\mathcal{S}} (s(x,y) - c(x,y))^2$$

where the summation takes place over the set $\mathcal{S}$ of all points in the representation for which there is a known stereo depth value $c(x, y)$. The effect of this objective function is to minimize a least-squares fit through the known points, scaled relative to the original minimization problem. The constant β is a scale parameter to be determined by the degree of desired fit. Note that the constraints have, in this case, been incorporated directly into the objective function. Hence, the objective function may be optimized as if it were an unconstrained function.

The translation of this problem into the image domain yields the following discrete version of the objective function:

$$\text{minimize} \quad \sum_{i=1}^{m-2}\sum_{j=0}^{m-1}\left(s_{(i-1,j)} - 2s_{(i,j)} + s_{(i+1,j)}\right)^2$$

$$+\sum_{i=0}^{m-1}\sum_{j=1}^{m-2}\left(s_{(i,j-1)} - 2s_{(i,j)} + s_{(i,j+1)}\right)^2$$

$$+2\sum_{i=0}^{m-2}\sum_{j=0}^{m-2}\left(s_{(i,j)} - s_{(i+1,j)} - s_{(i,j+1)} + s_{(i+1,j+1)}\right)^2$$

$$+\beta\sum_{\mathcal{S}}\left(s_{(i,j)} - c_{(i,j)}\right)^2.$$

It is now possible to consider applying the conjugate gradient method to this problem. Recall that this method, when applied to the quadratic case, is considered the minimization of

$$\frac{1}{2}\mathbf{s}^T\mathbf{Q}\mathbf{s} - \mathbf{b}^T\mathbf{s}.$$

In this case, the vector $\mathbf{b}$ is given by

$$\mathbf{b} = -2\beta\mathbf{c}$$

where $\mathbf{c}$ is a vector whose components are the known depth values $c_{(i,j)}$ if the corresponding grid point has such a known value, and 0 otherwise. The matrix $\mathbf{Q}$ is given by the discrete stencils outlined in the previous section, with an added diagonal factor of $\delta_{\mathcal{S}}$. One can then straightforwardly apply the conjugate gradient algorithm, with these forms for $\mathbf{b}$ and $\mathbf{Q}$.

6.5 Examples of Approximation

The conjugate gradient algorithm, applied to the surface approximation problem, is demonstrated by considering a series of examples, illustrated in Figures 12-16. These should be compared to Figures 8-11. As in the case of surface interpolation, the surface approximation algorithm has been applied to disparity values rather than depth information. Again, the general shape of the surface and the relative difference in positions of the surfaces have been preserved by the algorithm, although the exact surface shape has not been reconstructed.

The objective function of the conjugate gradient algorithm in this case contains two factors, the quadratic variation of the surface and a least-squares term embodying a type of "smoothness" requirement. The scalar constant β determines the relative strengths of these two factors. If we let β be very small, then the smoothness requirement essentially vanishes and we return to the case of surface interpolation, discussed previously. The conjugate gradient algorithm then becomes identical to the gradient projection algorithm. If we let β become very large, then the quadratic variation factor essentially vanishes and the algorithm reduces to a least-squares fitting of a plane to the known points. Clearly, we require a value of β intermediate to these extreme cases. The figures illustrate this tradeoff between the two factors, as β varies. To determine the optimal value for β, we require an estimate for the density of incorrect disparity values obtained by the stereo algorithm, so that a value for β may be chosen which smooths out the effect of these incorrect values, while not affecting the shape of the surface determined by minimizing the quadratic variation.

7. Analysis and Refinements

7.1 Discontinuities

One of the implicit assumptions of the interpolation algorithm is that the pieces of surface are in fact pieces of a single surface. Of course, this will frequently not be the case. In this section, we consider what modificiations are necessary in order to account for the existence of several surfaces within a scene. In particular, we address the issue of explicitly computing discontinuities in the surface representation, and the effects of explicit discontinuities on the form of the reconstructed surface.

One of the problems associated with the failure to make surface discontinuities explicit is that information about the shape of one surface affects the shape of an adjacent surface. This is illustrated in Figure 17. A set of known depth points is given in Figure 17(*a*). Intuitively, the most likely surface to fit through these points would be a pair of planes with a discontinuity in depth between them, shown in Figure 17(*b*). However, the requirement that a smooth surface fit through these points results in a warping and rippling of the surface that is undesirable, as shown in Figure 17(*c*). Thus, the lack of explicit discontinuities can affect the shapes of the interpolated surfaces in an unacceptable manner.

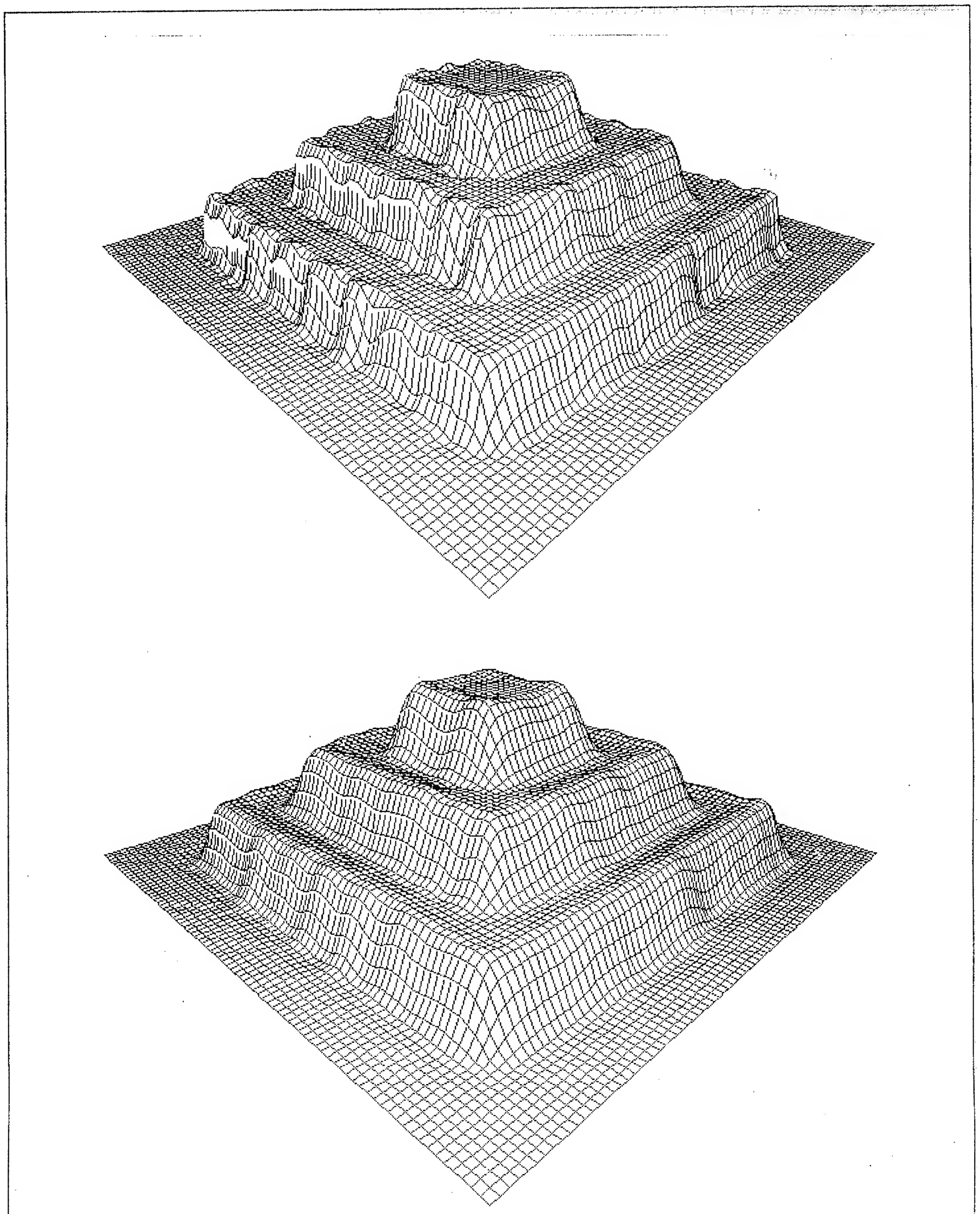

Figure 12. The Wedding Cake. The figures show surfaces obtained by approximating the surface using quadratic variation. The scalar constant relating the least squares term to the quadratic term is $\beta = 1$ in the top figure and $\beta = 0.1$ in the bottom figure.

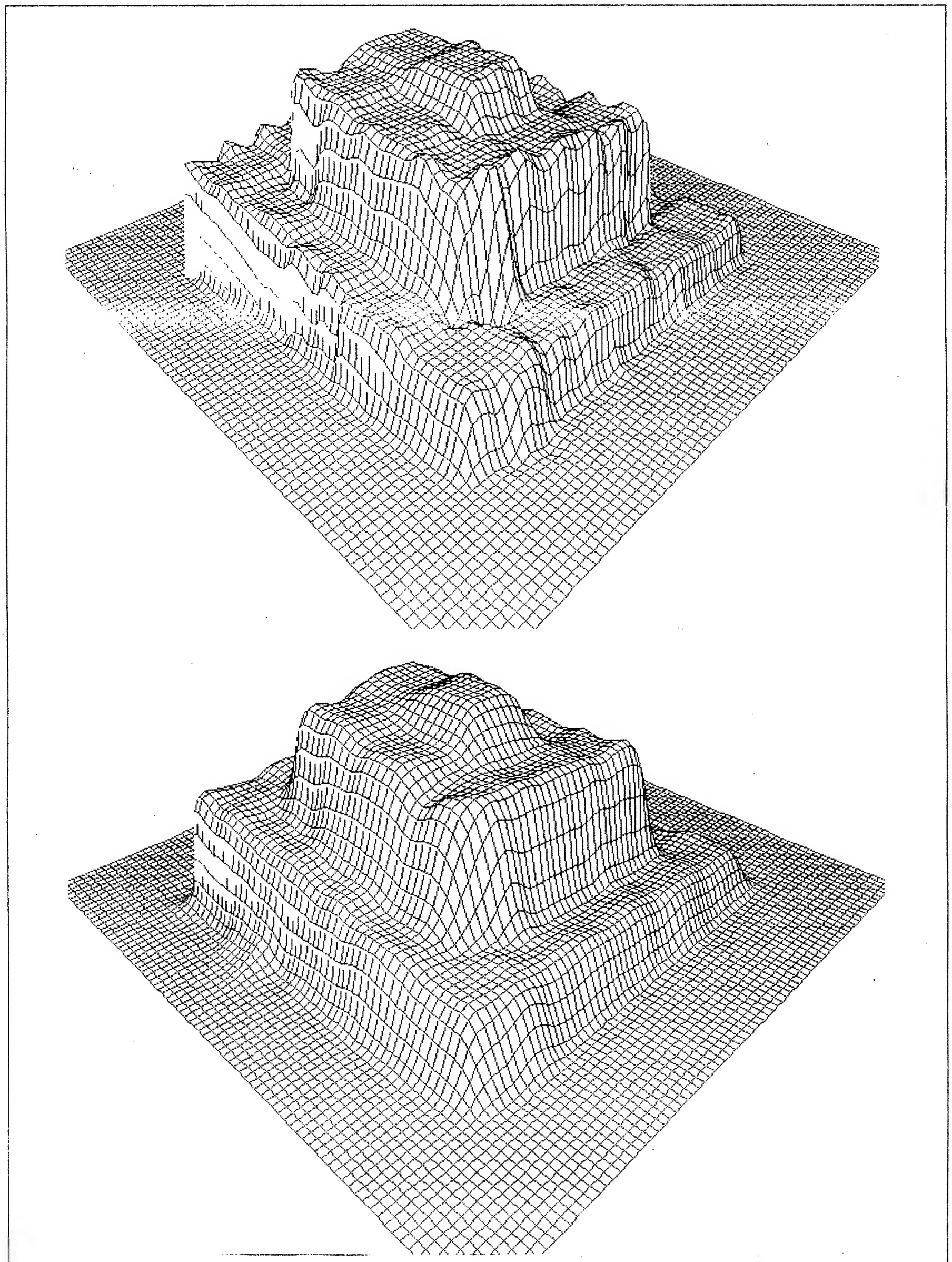

Figure 13. The Spiral Staircase. The figures show surfaces obtained by approximating the surface using quadratic variation. The scalar constant relating the least squares term to the quadratic term is $\beta = 1$ in the top figure and $\beta = 0.1$ in the bottom figure.

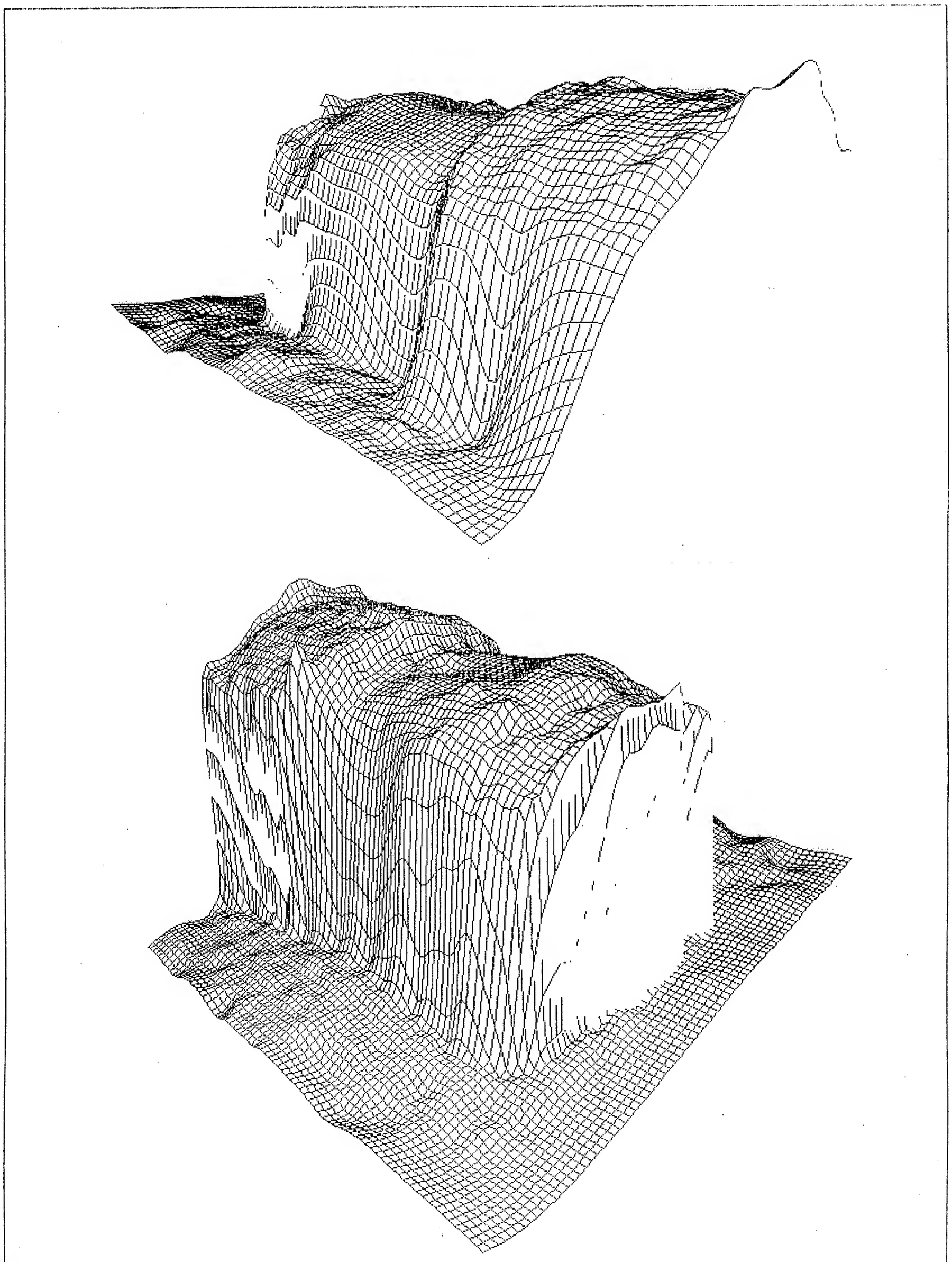

Figure 14. The Coffee Jar. The figures show two views of a surface obtained by approximating the surface using quadratic variation. The scalar constant relating the least squares term to the quadratic term is $\beta = 1$.

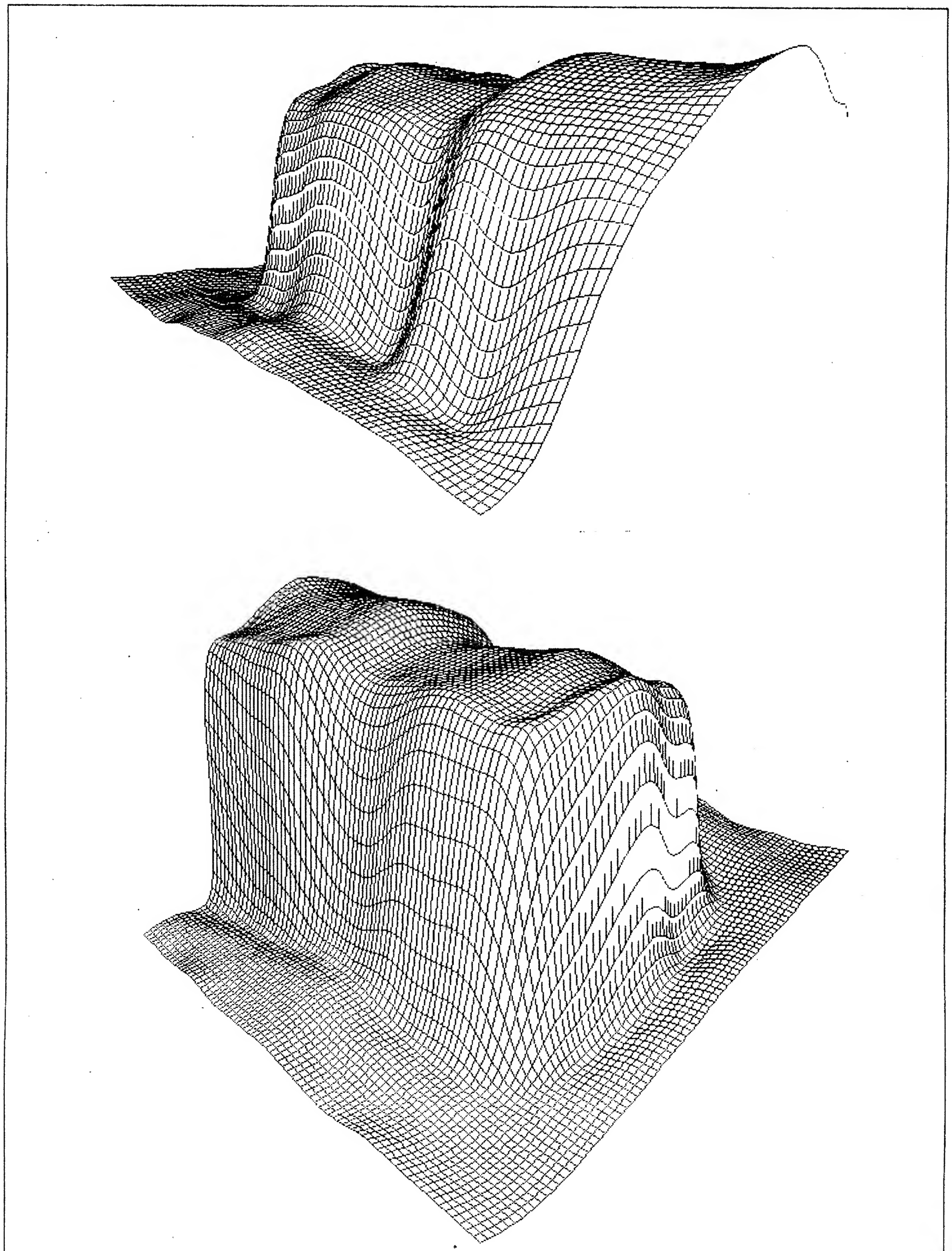

Figure 15. The Coffee Jar. The figures show two views of a surface obtained by approximating the surface using quadratic variation. The scalar constant relating the least squares term to the quadratic term is $\beta = 0.01$.

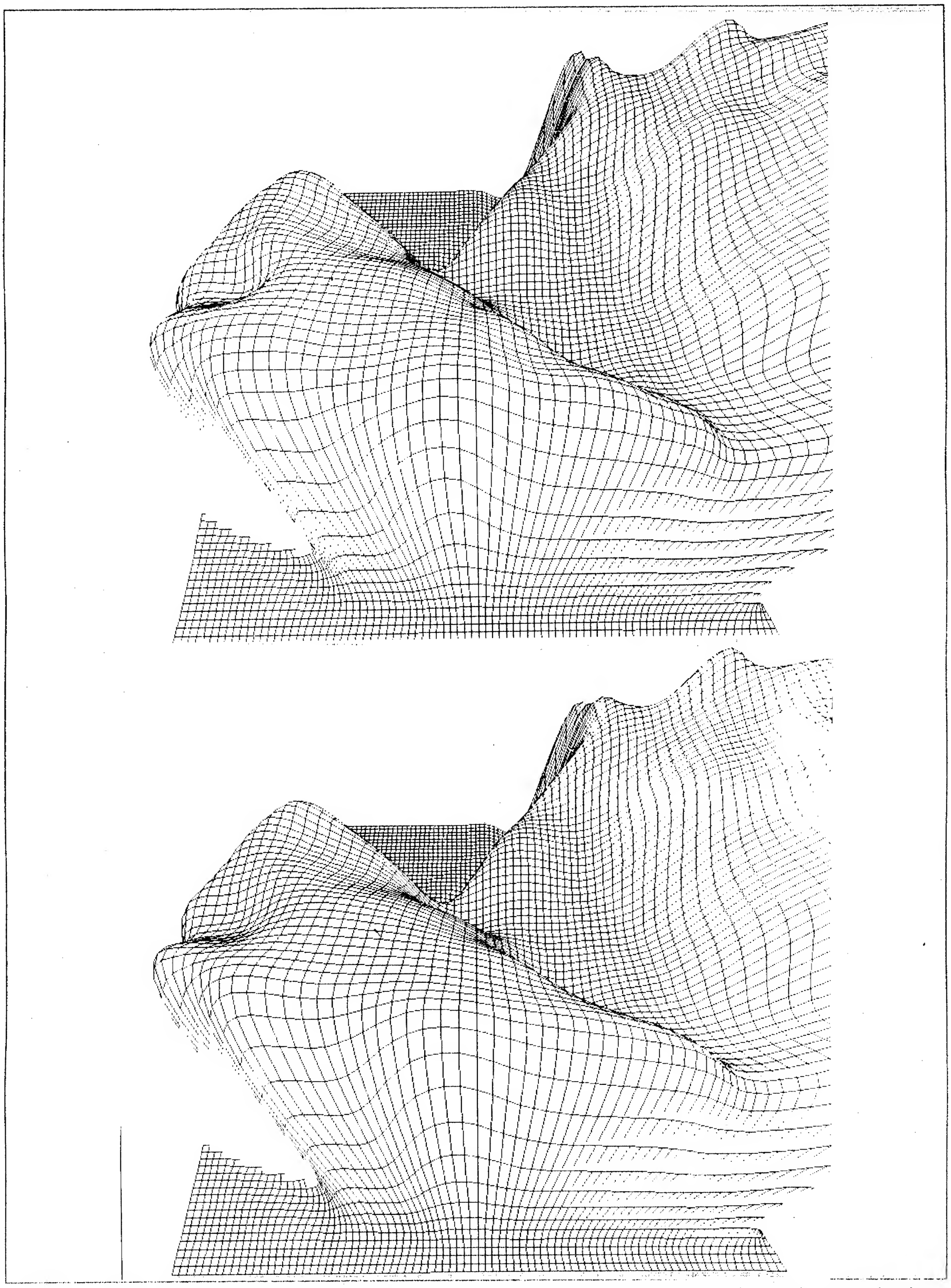

Figure 16. The Moore Sculpture. The figures show two approximations of the surface obtained by using quadratic variation. The scalar constant relating the least squares term to the quadratic term is $\beta = 0.5$ in the top figure and $\beta = 0.05$ in the bottom figure.

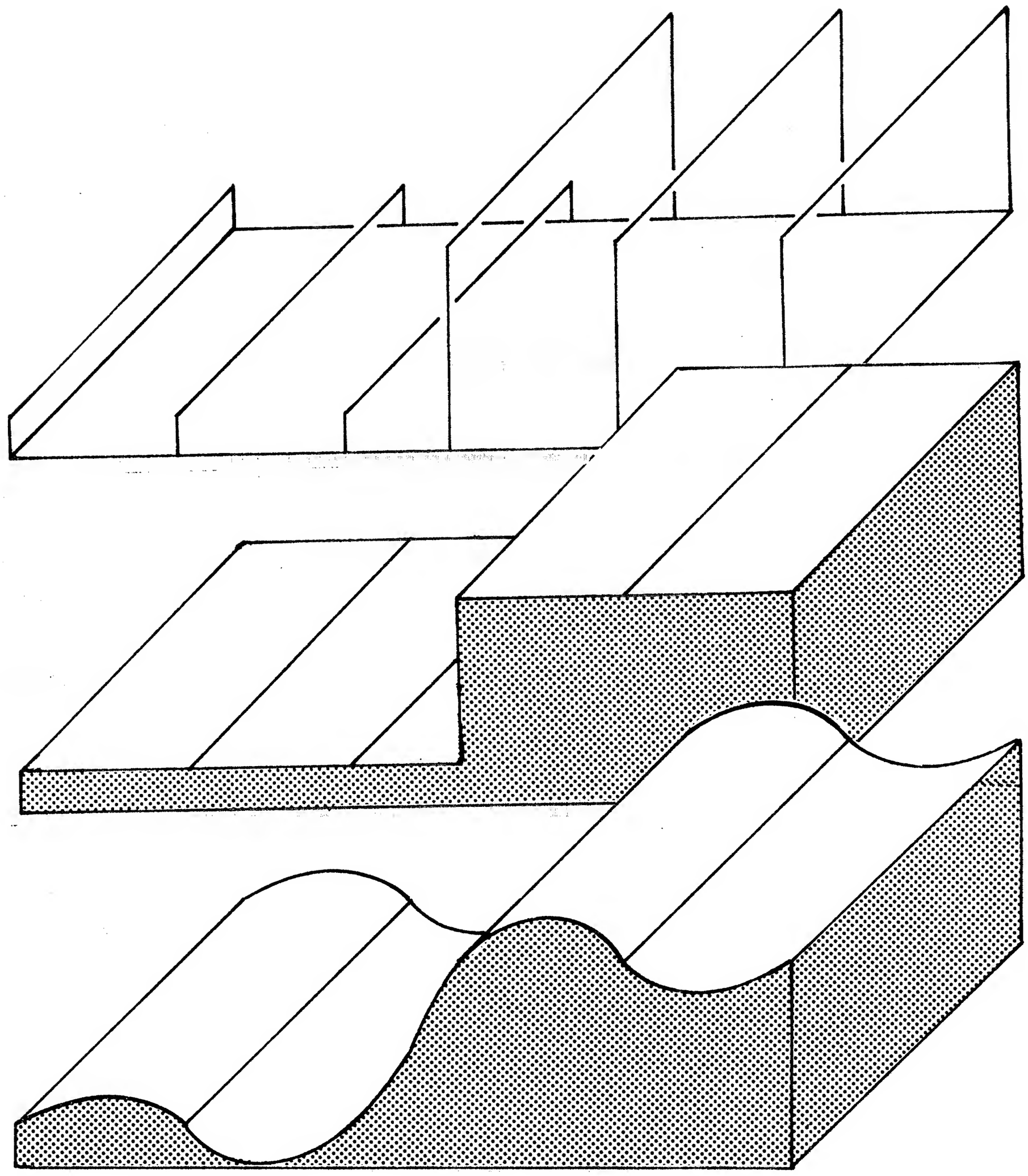

Figure 17. Discontinuities in the Surfaces. Figure (*a*) shows a set of known data points. Intuitively, the correct reconstructed surface would be a pair of planes, with a discontinuity between them, as shown in figure (*b*). If the interpolation algorithm attempts to reconstruct a surface through the boundary points, without a discontinuity, the result is as shown in figure (*c*). The sharp change in depth results in a rippling of the surface.

In order to make discontinuities explicit, there are several questions to ask about the process. How are the discontinuities detected? Where are they placed in the representation? When does the detection of discontinuities take place in the overall interpolation process? In the next few sections, we will discuss two possible methods for detecting the discontinuities, and their role in the overall interpolation.

7.1.1 Occlusions in the Stereo Algorithm

Consider the geometry indicated in Figure 18. There are regions of the left image which will not have a corresponding region in the right image, and vice versa. Consequently, any zero-crossings in this portion of one image will have no counterpart in the other image, and the stereo algorithm should not assign any match to such zero-crossings. Hence, one possible mechanism for detecting occlusions would be to search for portions of the image which contain unmatched zero-crossings. Then, the interpolation can be restricted to take place only over those sections of the image which are bounded by zero-crossings with known disparity values.

This method would detect the discontinuities before the interpolation, since it uses stereo information directly to locate the occlusions. A problem with the method is that it will not detect all discontinuities, only those in the horizontal direction. Discontinuities that occur in the vertical direction do not cause occlusions. Hence, any method for detecting discontinuities which relies only on the unmatched zero-crossings will be incomplete.

7.1.2 The Primal Sketch Revisited

An integral part of most computational theories, proposed as models of aspects of the human visual system, is the use of computational constraints based on assumptions about the physical world [Marr, 1976, 1980; Marr and Poggio, 1979; Marr and Hildreth, 1980; Ullman, 1979]. In some of the computational theories, the constraints are explicitly checked for validity within the algorithm (e.g. Ullman's rigidity constraint in recovering structure from motion). In others, the constraints are simply assumed to be true, and are not explicitly checked (e.g. Marr and Poggio's uniqueness constraint in stereopsis). Can any aspect of the surface consistency constraint be explicitly checked and used by the algorithm?

The basic notion of the surface consistency constraint is that the surface cannot undergo a radical change in shape without having an accompanying zero-crossing in the convolved image. Implicit

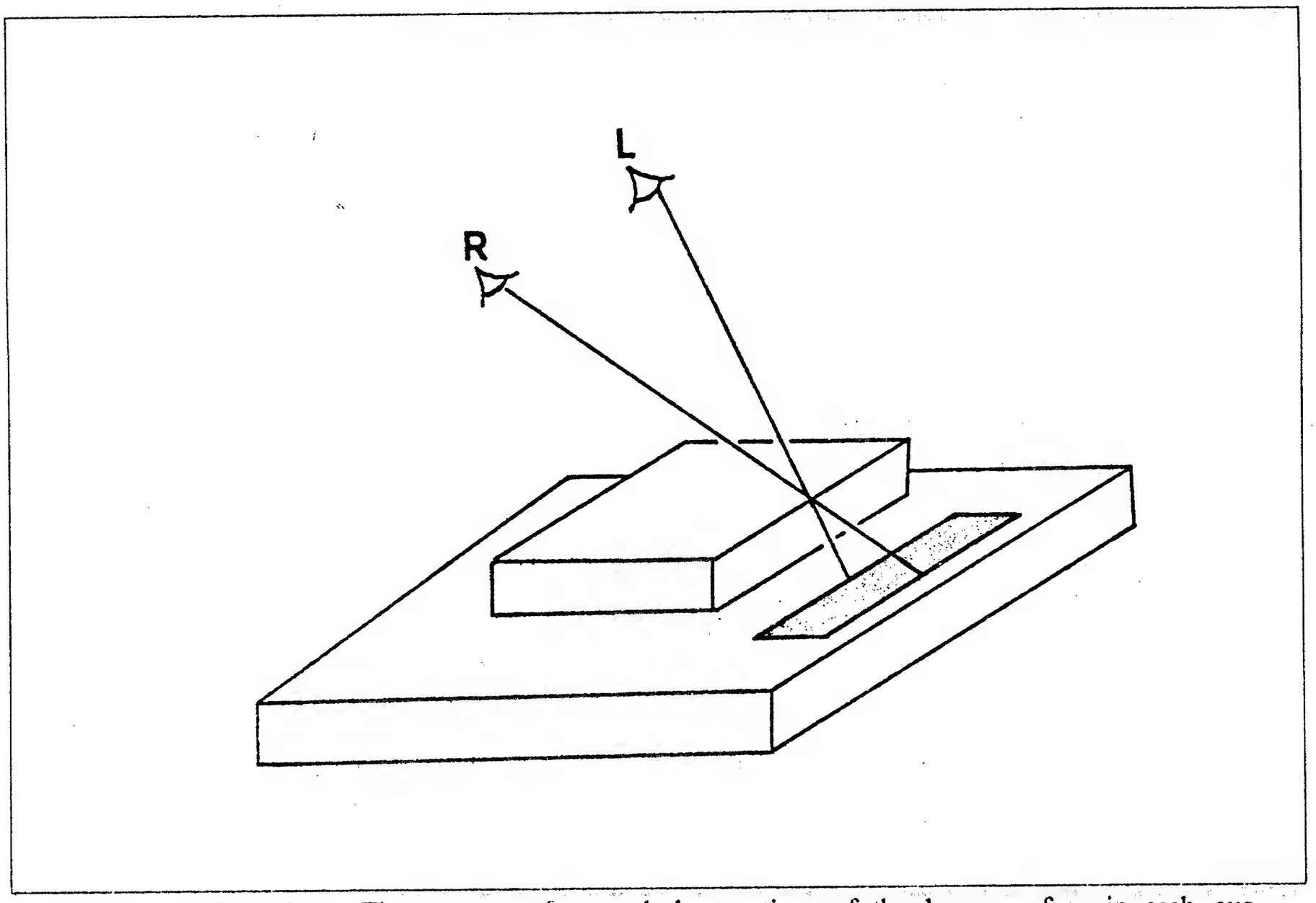

Figure 18. Occlusions. The upper surface occludes portions of the lower surface in each eye. These portions are different for the two eyes. The cross-hatched area of the lower surface indicates the region of the surface visible to the left eye, but not to the right.

in this constraint is the assumption that the portion of the image being examined in fact corresponds to a single object. Thus, one could propose that if the shape of the interpolated surface forces a zero-crossing in a location for which none exists in the Primal Sketch, then such a zero-crossing indicates a location at which the assumption of a single object is violated. Such zero-crossings could then be taken as indicative of a surface discontinuity.

Perhaps the simplest method of detecting such discontintuities is again to use ideas inherent in the Primal Sketch. Recall that the Primal Sketch created descriptions of points in the image associated with inflections in intensity, for a range of resolutions. Since the image intensities may be considered as a type of three-dimensional surface, the Primal Sketch operators essentially detect discontinuities in the image intensities for a range of resolutions. Thus, one could apply the same type of analysis to the detection of surface discontinuities, where now the surface on which the operators apply is the reconstructed depth surface, rather than the intensity surface.

It is worth noting that not only should the operators be of the form used in the extraction of the Primal Sketch, but that it may also be useful to use a range of operators, as in the Primal Sketch. One reason for using multiple zero-crossing detectors was that surface changes, and hence intensity changes, could take place over a wide range of scales. This is still true in the case of surface descriptions, such as have been constructed for the coffee jar or the wedding cake. Thus, surface discontinuities corresponding to occluding edges will frequently tend to correspond to large surface changes, while internal surface discontinuities, due to a warping of the surface, will tend to correspond to small surface changes. By using a range of $\nabla^2 G$ operators, one can extract both occluding contour discontinuities, as well as ripples or warpings of the surface itself.

Note that this method requires that the surface interpolation already take place, before it can be applied. Since one of the general requirements on an algorithm is that it be rapid, we must consider the consequence of detecting discontinuities after the interpolation of the surfaces. There are two main reasons for the explicit detection of discontinuities. One is that such an explicit representation of this information will allow higher level processes, such as recognition, or extraction of axes for three-dimensional shape analysis, to operate more easily, since the process serves to make implicit information explicit. However, a second reason is to create more accurate surface representations, by removing the type of effect illustrated in Figure 17(*c*). If the process used to isolate discontinuities takes place after interpolation, and if the interpolation process requires the discontinuities to improve the interpolated surface approximation, one must propose an interpolater which passes over the surface information twice; first to produce an initial description, and second to refine the description after the detection of discontinuities. One must then question whether such a two pass process will affect our constraint of rapid algorithms. Fortunately, the answer is no, since the surface approximation obtained without explicitly accounting for the discontinuities is very close to the limiting surface except in the areas of the discontinuities (that is, any effects of the discontinuities are quickly damped out as one moves across the surface). Thus, the initial starting position for the second pass of the interpolation algorithm is very close to the limiting surface, and only a few iterations will be needed to refine the surface approximation.

7.1.3 Interpolation Over Occluded Regions

Even though occluded regions of the image can only be viewed from one eye, the human system still associates a depth value with these regions. This has an interesting implication for the interpolation algorithm. For most occluded regions, the only depth information available is at the edges of the occluded region. Psychophysical experiments have shown that the occluded region is always perceived at the depth of the lower surface. Thus, in Figure 18, the occluded region would be perceived at the level of the lower surface. Note that this is consistent with the physics of the situation, since if the occluded region were perceived at the level of the upper surface, then it should in fact be visible to the right eye, and this is not the case.

This observation suggests that when an occlusion is detected, it is explicitly located along the occluding boundary corresponding to the edge of the nearer object. This allows the occluded region itself to be associated with the lower surface, and the interpolation algorithm will fill in surface values for the occluded region from this lower surface.

This raises an interesting psychophysical prediction. The psychophysical literature has examined the case of planar surfaces and their occlusions, as in Figure 18. If the interpolation method developed here is given an explicit discontinuity along one edge of the occluded region, it will correctly fill in the region as an extension of the lower plane. Of interest is the case in which the occluded region is not planar. For example, consider a cylindrical object. If the interpolation algorithm is given this type of input, it will fill in the occluded portions of the surfaces as a smooth continuation of the curved cylinder. If the interpolation algorithm correctly models interpolation by the human visual system, then this predicts that the surface perception for human observers in this situation should also be that of a smooth cylinder. While informal experiments indicate that this is true, the prediction has not yet been rigourously tested psychophysically.

7.2 Noise Removal

Although in general the Marr-Poggio stereo algorithm is very good at matching zero-crossings correctly (especially for random dot patterns), incorrect disparity values may sometimes be assigned to regions of the image. These incorrect values can be considered as noise superimposed on the correct surface. Since the surface interpolator explicitly attempts to fit a surface through all the disparity points, such noise points can affect the shape of the surface approximation. Indeed, the effect of these

noise points can spread over a noticeable portion of the surface, before the nearby disparity values can damp out its effect. Thus, it would be preferable to remove these noise points, or at least neutralize their effect on the approximated surface shape. One possibility is that if a two pass interpolator is used, as suggested in the previous section, the detection of surface discontinuities will isolate such noise points from the rest of the surface, and the second pass of the interpolator will adjust the surface approximation to remove the influence of the noise points on the first pass approximation. Certainly this will be true for noise points with disparity values far removed from the correct values. For noise points whose disparity values are only slightly different from the correct surface disparities, the difference does not really matter. However, the final result would be that the noise points, while being isolated from the rest of the correct surface, would still remain in the final surface description. It would be preferable to completely remove such points.

Is it possible to identify and remove noise points from the disparity map? If the noise points are isolated spatially, then it is possible to identify them as undesirable. This follows from the form of the primal sketch operators. The case to consider is that in which one must distinguish between a set of noise points in a disparity map and a small object separated in depth from the rest of the scene. For the small object, the size of the zero-crossing contour is limited by the size of the available operator, and hence there is a minimum size of zero-crossing contour which the operator will yield about the object. If the number of zero-crossing points which differ significantly from their neighbors is less than this minimum, one may conclude that the points are noise, and thus remove them. This will result in an improved surface approximation.

7.3 Acuity

It can be seen from the example of the interpolated coffee jar in Figure 10, that the interpolated surface contains a bumpy quality which clearly is not consistent with the original object. How can this be explained? The effect occurs in part because the disparity values are specified only to within a pixel. This yields a fairly coarse disparity map which results in the bumps observed in the interpolated coffee jar of Figures 14 and 15. Hence, one method of removing the bumps would be to improve the accuracy of the disparities obtained by the algorithm. Note that some improvement in disparity accuracy is necessary if the algorithm is to be consistent with the human system. If we roughly equate pixels with receptors, then a pixel corresponds to roughly 27 seconds of arc. The implementation of

the stereo algorithm computed disparity to within a pixel, while humans are capable of stereo acuity to a resolution of 2 — 10 seconds [Howard, 1919; Woodburne, 1934; Berry, 1948; Tyler, 1977].

In order to account for finer disparity values, it is necessary to localize the zero-crossing to a better accuracy than has been done so far. Since the convolution values are only specified at each pixel, one method for more accurately specifying the zero-crossing positions is to interpolate between the known convolution values [Crick, Marr and Poggio 1980, Marr, Poggio and Hildreth 1979, Hildreth 1980]. Perhaps the simplest method is to rely on the observation of Hildreth that for most cases, even a simple linear interpolation will give extremely accurate localization of the zero-crossings. The addition of finer resolution depth information may improve the performance of the algorithm.

This example also raises a question of scale. Depending on the application of the surface specification, different amounts of resolution may be required. For example, if the ultimate goal of the surface specification is to obtain a rough idea of the position and shape of the surfaces in a scene, the spatial resolution at which surface information must be made explicit may not be critical. In this case, the known data from the stereo algorithm may be sampled at a coarser resolution, before the interpolation takes place. This should result in a smoother surface approximation. Further, although the reconstructed surface is less exact in terms of fine variation of the surface shape, the overall shape of the bottle is still preserved in this interpolation.

7.4 Psychophysics

We close by listing a series of psychophysical questions of relevance to the interpolation process.

(1) What is the form of the surface perceived in occluded regions? In particular, the minimization of quadratic variation suggests that if a portion of a curved object is occluded, then the surface in the occluded region should also be curved, and should minimize the quadratic variation across that region.

(2) Figure 17 suggests that if discontinuities are not explicitly demarked in the interpolation process, a warping of the reconstructed surface (similar to Gibb's phenomena) will result. While, in principle, such ripples in the surface are undesirable, it is worth asking whether the human system specifically accounts for discontinuities before interpolation occurs. This may be rephrased by asking whether in stereoscopic situations similar to Figure 17, we perceive a Mach band-like warping of the surface in depth?

(3) We have suggested that there are several possible functionals which could be used to determine the most consistent surface. Based on algorithmic and mathematical arguments, we choose the quadratic variation. Can we test the shape of the reconstructed surface psychophysically? In particular, can we distinguish psychophysically between the minimum surface under quadratic variation and the minimum surface under some other functional, such as the square Laplacian? Is the reconstructed surface psychophysically consistent with the surface computed by quadratic variation?

(4) What is the spatial resolution of the reconstructed surface? That is, what is the spacing of the grid upon which the values of the reconstructed surface are computed?

The answers to these questions will help verify or correct the theory of visual surface interpolation developed in this paper.

8. Summary

Computational theories of motion perception [Ullman, 1979] and stereo vision [Marr and Poggio, 1979] can only specify the computation of three-dimensional surface information at special points in the image. In order to account for the visual perception of complete surfaces, we have developed a computational theory of the interpolation of surfaces from visual information.

The problem is constrained by the fact that the surface must agree with the information from stereo or motion correspondence, and **not** vary radically between these points. In Grimson [1981c], an explicit form of this *surface consistency constraint* is derived from the image intensity equation [Horn, 1975]. The main point of the surface consistency constraint is that it requires the interpolated surface to vary as little as possible.

To determine which of two possible surfaces is more consistent with the surface consistency constraint, one must be able to compare the two surfaces. To do this, a functional from the space of functions to the real numbers is required, where the functional should measure some function of the variation in the surface. In this way, the surface most consistent with the visual information will be that which minimizes the functional. To ensure that the functional has a unique minimal surface, conditions on the form of the functional are derived. In particular, if the functional is a complete semi-norm which satisfies the parallelogram law, or the space of functions is a semi-Hilbert space and

the functional is a semi-inner product, then there is a unique (to within an element of the null space of the functional) surface which is most consistent with the visual information.

It can be shown, based on the above conditions plus a condition of rotational symmetry, that there is a vector space of possible functionals which measure surface consistency, this vector space being spanned by the functional of quadratic variation and the functional of square Laplacian (Brady and Horn, 1981). Arguments based on the null spaces of the respective functionals were used to justify the choice of the quadratic variation as the optimal functional.

Algorithms for computing the surface which minimizes quadratic variation in the case of exact surface interpolation and in the case of surface approximation were outlined and illustrated on a series of synthetic and actual surface interpolation examples.

9. Acknowledgements

The author wishes to express his gratitude for many useful comments and discussions to David Marr, Tommy Poggio, Shimon Ullman, Berthold Horn, Mike Brady, Whitman Richards, Tomas Lozano-Perez, Marilyn Matz and Ellen Hildreth.

10. References

Abramowitz, M. and Stegun, I.A. *Handbook of Mathematical Functions*, Dover Publications, Inc., New York, 1965.

Barrow, H.G. and Tenenbaum, J.M. "Interpreting line drawings as three-dimensional surfaces," *Artificial Intelligence (Special Issue on Computer Vision)* 17 (1981).

Berry, R.N. "Quantitative relations among vernier, real depth, and stereoscopic depth acuities," *J. Exp. Psychol.* 38 (1948), 708-721.

Brady, J.M and Horn, B.K.P "Rotationally symmetric operators for surface interpolation," (1981), to appear.

Crick, F.H.C., Marr, D. and Poggio, T. "An information processing approach to understanding the visual cortex," *in The Cerebral Cortex, N.R.P.* (1980).

Courant, R. and Hilbert, D. *Methods of Mathematical Physics, Volume I*, Interscience Publishers, Inc., New York, 1953.

Davis, L. "A survey of edge detection techniques," *Computer Graphics and Image Processing* **4** (1975), 248-270.

Duchon, J. Fonctions-spline du type plaque mince en dimension 2, Univ. of Grenoble, Rpt. 231, 1975.

Duchon, J. Fonctions-spline a energie invariante par rotation, Univ. of Grenoble, Rpt. 27, 1976.

Forsyth, A.R. *Calculus of Variations*, Dover Publications, Inc., New York, 1960.

Grimson, W.E.L. Computing shape using a theory of human stereo vision, Ph.D. Thesis, Department of Mathematics, Massachusetts Institute of Technology, 1980.

Grimson, W.E.L. "A computer implementation of a theory of human stereo vision," *Phil. Trans. Royal Society of London, B.* **292** (1981a), 217-253.

Grimson, W.E.L. *From Images to Surfaces: A computational study of the human early visual system*, MIT Press, Cambridge, Mass., 1981b.

Grimson, W.E.L. "The implicit constraints of the primal sketch," (1981c), to appear.

Grimson, W.E.L. and Marr, D. "A computer implementation of a theory of human stereo vision," *Proceedings: Image Understanding Workshop*, Palo Alto, Cal., 1979, 41-47.

Hildreth, E.C. Implementation of a theory of edge detection, M. Sc. Thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 1980.

Horn, B.K.P. Shape from shading: a method for obtaining the shape of a smooth spaque object from one view, MIT Project MAC Technical Report, MAC TR-79, 1970.

Horn, B.K.P. "Obtaining shape from shading information," *The Psychology of Computer Vision, P.H. Winston (ed), McGraw-Hill* (1975), 115-155.

Horn, B.K.P. "Understanding image irradiances," *Artificial Intelligence* **8** (1977), 201-231.

Horn, B.K.P. and Bachman, B.L. Using synthetic images to register real images with surface models, MIT Artificial Intelligence Laboratory, Memo 437, 1977.

Horn, B.K.P. and Sjoberg, R.W. "Calculating the reflectance map," *Applied Optics* **18** (1979), 1770-1779.

Howard, J.H. "A test for the judgement of distance," *Am. J. Ophthal.* **2** (1919), 656-675.

Huffman, D.A. "Impossible objects as nonsense sentences," *Machine Intelligence 6, B. Meltzer and D. Michie (eds), Edinburgh University Press* (1971), 295-393.

Hummel, R.A. and Zucker, S.W. On the foundations of relaxation labeling processes, Computer Vision and Graphics Lab., McGill University, TR-80-7, 1980.

Julesz, B. "Binocular depth perception of computer-generated patterns," *Bell System Tech. J.* **39** (1960), 1125-1162.

Luenberger, D.G. *Introduction to Linear and Nonlinear Programming* , Addison-Wesley Publishing Co., Reading, Mass., 1973.

Mackworth, A.K. "Interpreting pictures of polyhedral scenes," *Artificial Intelligence* **4**, 2, 121-137.

Marr, D. "Early processing of visual information," *Philosophical Transactions of the Royal Society of London* **275**, 942 (1976), 483-534.

Marr, D. "Representing visual information," *AAAS 143rd Annual Meeting, Symposium on Some Mathematical Questions in Biology, February 1977. Published in Lectures in the Life Sciences* **10** (1978). 101-180.

Marr, D. *VISION: A computational investigation in the human representation and processing of visual information* , W.J. Freeman, San Francisco, 1981.

Marr, D. and Hildreth, E.C. "Theory of edge detection," *Proc. R. Soc. Lond. B* **207** (1980), 187-217.

Marr, D. and Nishihara, H.K. "Representation and recognition of the spatial organization of three-dimensional shapes," *Proc. R. Soc. Lond. B.* (1978).

Marr, D. and Poggio, T. "From understanding computation to understanding neural circuitry," *Neuroscience Research Program Bulletin* **15**, 3 (1977), 470-488.

Marr, D. and Poggio, T. "A theory of human stereo vision," *Proc. R. Soc. Lond. B* **204** (1979), 301-328.

Marr, D., Poggio, T. and Hildreth, E. "The smallest channel in early human vision," *JOSA* **70**, 7 (1979), 868-870.

Mayhew, J.E.W. and Frisby, J.P. "Psychophysical and computational studies towards a theory of human stereopsis," *Artificial Intelligence (Special Issue on Computer Vision)* **17** (1981).

O'Brien, B. "Vision and resolution in the central retina," *J. Opt. Soc. Am.* **41** (1951), 882-894.

Pratt, W. *Digitial Image Processing* , John Wiley and Sons, New York, 1978.

Rosenfeld, A. and Kak, A. *Digital Picture Processing* , Academic Press, New York, 1976.

Rudin, W. *Functional Analysis* , McGraw-Hill Book Company, New York, 1973.

Schumaker, L. L. "Fitting surfaces to scattered data," *Approximation Theory II, ed. G.G. Lorentz, C.K. Chui, L.L. Schumaker, Academic Press* (1976), 203-268.

Stevens, K.A. Surface perception from local analysis of texture and contour, MIT AI Laboratory , TR 512, 1979.

Tyler, C.W. "Spatial limitations of human stereoscopic vision," *SPIE* **120** (1977).

Ullman, S. *The Interpretation of Visual Motion* , MIT Press, Cambridge, Mass., 1979a.

Ullman, S. "Relaxation and constrained optimization by local processes," *Computer Graphics and Image Processing* **10** (1979b), 115-125.

Witkin, A.P. Shape from contour, Ph. D. Thesis, Department of Psychology, Massachusetts Institute of Technology, 1980.

Woodburne, L.S. "The effect of a constant visual angle upon the binocular discrimination of depth differences," *Am. J. Psychol.* 46 (1934), 273-286.

Woodham, R.J. Reflectance map techniques for analyzing surface defects in metal castings, MIT AI Lab, TR 457, 1978.